247
1.9.70

Child Development

VOLUME 32

MARCH, 1961 *to* DECEMBER, 1961

Editor

William E. Martin
Purdue University

Consulting Editors

Donald M. Baer
University of Washington

Emanuel K. Beller
Child Development Center

Alfred Castaneda
University of Texas

Walter Emmerich
University of Colorado

Jerome Kagan
Fels Research Institute

Irvin A. Kraft
Houston State Psychiatric Institute

Lewis P. Lipsitt
Brown University

Norman Livson
University of California

James M. Tanner
University of London

Richard H. Walters
University of Toronto

Managing Editor

Kate Hoffman

BUREAU OF EDNL. & PSYL. RESEARCH
LIBRARY

CHILD DEVELOPMENT PUBLICATIONS

Society for Research in Child Development, Inc.
Purdue University, Lafayette, Indiana

1961

Bureau Edn. Research
DAVID COLLEGE
Dated 1·9·70
Acc. No. 9247

CONTENTS

Volume 32, 1961

Acker, Mary. (*see* McReynolds, Paul) 393
Allinsmith, Wesley. (*see* Burton, Roger V.) 689
Asher, John William. Comment on "The Relationship Between Rigidity-Flexibility in Children and Their Parents" 607

Baer, Donald M. Effect of Withdrawal of Positive Reinforcement on an Extinguishing Response in Young Children 67
Baker, Charles T. (*see* Lansky, Leonard M.) 45
Baker, John W., & Holzworth, Annette. Social Histories of Successful and Unsuccessful Children 135
Baldwin, Alfred L. (*see* Paivio, Allan) 721
Banta, Thomas J. (*see* Eron, Leonard D.) 457
Barnard, James W., Zimbardo, Philip G., & Sarason, Seymour B. Anxiety and Verbal Behavior in Children 379
Bartram, John B. (*see* Zuk, G. H.) 525
Becker, Wesley C. (*see* Peterson, Donald R.) 151
Bennett, Hannah Lou. (*see* McCandless, Boyd R.) 511
Berger, Seymour M. (*see* Paivio, Allan) 721
Bijou, Sidney W. Discrimination Performance as a Baseline for Individual Analysis of Young Children 163
Bilous, Carolyn B. (*see* McCandless, Boyd R.) 511
Borstelmann, L. J. Sex of Experimenter and Sex-Typed Behavior of Young Children 519
Bronson, Wanda C. (*see* Livson, Norman) 75
Brown, Ann Wilson, & Hunt, Raymond G. Relations between Nursery School Attendance and Teachers' Ratings of Some Aspects of Children's Adjustment in Kindergarten 585
Brown, Kenneth B. (*see* Ourth, Lynn) 287
Burchinal, Lee G. (*see* Gardner, D. Bruce) 225
Burton, Roger V., Maccoby, Eleanor E., & Allinsmith, Wesley. Antecedents of Resistance to Temptation in Four-Year-Old Children 689

Castaneda, Alfred, Fahel, Leila Snyder, & Odom, Richard. Associative Characteristics of Sixty-Three Adjectives and Their Relation to Verbal Paired-Associate Learning in Children 297
Cox, F. N., & Leaper, Patricia M. Assessing Some Aspects of the Parent-Child Relationship 637 ✓
Craig, Grace J. (*see* Myers, Nancy A.) 765
Crandall, Vaughn J., & Preston, Anne. Verbally Expressed Needs and Overt Maternal Behaviors 261
Crandall, Vaughn J. (*see also* Lansky, Leonard M.) 45

Davidson, Kenneth S., & Sarason, Seymour B. Test Anxiety and Classroom Observations 199
Dentler, Robert A., & Hutchinson, J. Glenn. Socioeconomic versus Family Membership Status as Sources of Family Attitude Consensus 249
Dittman, Allen T., & Goodrich, D. Wells. A Comparison of Social Behavior in Normal and Hyperaggressive Preadolescent Boys 315
Downing, Robert W., Moed, George, & Wight, Byron W. Studies of Disability: A Technique for Measurement of Psychological Effects 561
Draguns, Juris G., & Multari, Giuseppe. Recognition of Perceptually Ambiguous Stimuli in Grade School Children 541

Elkind, David. Quantity Conceptions in Junior and Senior High School Students 551
Emmerich, Walter. Family Role Concepts of Children Ages Six to Ten 609
Eron, Leonard D., Banta, Thomas J, Walder, Leopold O., & Laulicht, Jerome H. Comparison of Data Obtained from Mothers and Fathers on Childrearing Practices and Their Relation to Child Aggression 457
Escalona, Sibylle K., & Moriarty, Alice. Prediction of Schoolage Intelligence from Infant Tests 597
Estes, Betsy Worth. Judgment of Size in Relation to Geometric Shape 277

Fahel, Leila Snyder. (*see* Castaneda, Alfred) 297
Ferster, C. B. Positive Reinforcement and Behavioral Deficits of Autistic Children 437
Fort, Jane G. Secondary Reinforcement with Preschool Children 755

Gallistel, Elizabeth. (*see* Werner, Emmy) 255
Gardner, D. Bruce, Hawkes, Glenn R., & Burchinal, Lee G. Noncontinuous Mothering in Infancy and Development in Later Childhood 225
Goodrich, D. Wells. (*see* Dittmann, Allen T.) 315
Grinder, Robert E. New Techniques for Research in Children's Temptation Behavior 679
Gump, Paul V., & Kounin, Jacob S. Milieu Influences in Children's Concepts of Misconduct 711

Hawkes, Glenn R. (*see* Gardner, D. Bruce) 225
Hellmer, Leo A. (*see* Peterson, Donald R.) 151
Hendry, Louise S. (*see* Kessen, William) 95
Hoffman, Lois Wladis. Effects of Maternal Employment on the Child 187
Holloway, Harold D. Normative Data on the Children's Manifest Anxiety Scale at the Rural Third Grade Level 129
Holton, Ruth B. Amplitude of an Instrumental Response Following the Cessation of Reward 107
Holzworth, Annette. (*see* Baker, John W.) 135
Horowitz, Frances Degen. Latency of Sociometric Choice among Preschool Children 235
Hosken, Barbara. (*see* Kagan, Jerome) 625
Hunt, Raymond G. (*see* Brown, Ann Wilson) 585
Hutchinson, J. Glenn. (*see* Dentler, Robert A.) 249

Jeffrey, Wendell E. Variables in Early Discrimination Learning: III. Simultaneous vs. Successive Stimulus Presentation 305
Johnson, Carol. (*see* Johnson, Ronald C.) 271
Johnson, Ronald C., Johnson, Carol, & Martin, Lea. Authoritarianism, Occupation, and Sex Role Differentiation of Children 271

Kagan, Jerome, Hosken, Barbara, & Watson, Sara. Child's Symbolic Conceptualization of Parents 625
Kagan, Jerome, & Lemkin, Judith. Form, Color, and Size in Children's Conceptual Behavior 25
Kagan, Jerome. (*see also* Lansky, Leonard M.) 45
Keislar, Evan R. Experimental Development of "Like" and "Dislike" of Others among Adolescent Girls 59
Kessen, Marion Lord. (*see* Kessen, William) 779
Kessen, William, Hendry, Louise S., & Leutzendorff, Anne-Marie. Measurement of Movement in the Human Newborn: A New Technique 95
Kessen, William, & Kessen, Marion Lord. Behavior of Young Children in a Two-Choice Guessing Problem 779
Kessen, William, Williams, E. Jane, & Williams, Joanna P. Selection and Test of Response Measures in the Study of the Human Newborn 7
Kessen, William. (*see also* Williams, Joanna P.) 243
Kling, Frederick. (*see* Zuk, G. H.) 525
Kounin, Jacob S. (*see* Gump, Paul V.) 711
Krogman, Wilton Marion. A Human Biologist Looks at the Contemporary Scene 417

Lansky, Leonard M., Crandall, Vaughn J., Kagan, Jerome, & Baker, Charles T. Sex Differences in Aggression and Its Correlates in Middle-Class Adolescents ... 45
Laulicht, Jerome H. (*see* Eron, Leonard D.) ... 457
Leaper, Patricia M. (*see* Cox, F. N.) ... 637
Leiman, Alan H., Myers, Jerome L., & Myers, Nancy A. Secondary Reinforcement in a Discrimination Problem with Children ... 349
Lemkin, Judith. (*see* Kagan, Jerome) ... 25
Leutzendorff, Anne-Marie. (*see* Kessen, William) ... 95
Levy, Leon H. (*see* Matthews, Charles G.) ... 577
Lipsitt, Lewis P. Simultaneous and Successive Discrimination Learning in Children ... 337
Livson, Norman, & Bronson, Wanda C. An Exploration of Patterns of Impulse Control in Early Adolescence ... 75
Lövaas, O. Ivar. Effect of Exposure to Symbolic Aggression on Aggressive Behavior ... 37
Lövaas, O. Ivar. Interaction between Verbal and Nonverbal Behavior ... 329
Luria, Zella. (*see* Peterson, Donald R.) ... 151

McCandless, Boyd R., Bilous, Carolyn B., & Bennett, Hannah Lou. Peer Popularity and Dependence on Adults in Preschool-Age Socialization ... 511
Maccoby, Eleanor E. (*see* Burton, Roger V.) ... 689
McCord, Joan, & McCord, William. Cultural Stereotypes and the Validity of Interviews for Research in Child Development ... 171
McCord, William. (*see* McCord, Joan) ... 171
McCullers, John C. Effects of Associative Strength, Grade Level, and Interpair Interval in Verbal Paired-Associate Learning ... 773
McReynolds, Paul, Acker, Mary, & Pietila, Caryl. Relation of Object Curiosity to Psychological Adjustment in Children ... 393
Martin, Lea. (*see* Johnson, Ronald C.) ... 271
Matthews, Charles G., & Levy, Leon H. Response Sets and Manifest Anxiety Scores in a Retarded Population ... 577
Medinnus, Gene R. Q-Sort Descriptions of Five-Year-Old Children by Their Parents ... 473
Meyer, William J., & Seidman, Stanley B. Relative Effectiveness of Different Reinforcement Combinations on Concept Learning of Children at Two Developmental Levels ... 117
Miller, Ralph L. (*see* Zuk, G. H.) ... 525
Moed, George. (*see* Downing, Robert W.) ... 561
Moriarty, Alice. (*see* Escalona, Sibylle K.) ... 597
Morrow, William R., & Wilson, Robert C. Family Relations of Bright High-Achieving and Under-Achieving High School Boys ... 501
Multari, Giuseppe. (*see* Draguns, Juris G.) ... 541
Myers, Jerome L. (*see* Leiman, Alan H.) ... 349
Myers, Jerome L. (*see also* Myers, Nancy A.) ... 765
Myers, Nancy A., Craig, Grace J., & Myers, Jerome L. Secondary Reinforcement as a Function of the Number of Reinforced Trials ... 765
Myers, Nancy A. (*see also* Leiman, Alan H.) ... 349

Odom, Richard. (*see* Castaneda, Alfred) ... 297
Odom, Richard. (*see also* Stevenson, Harold W.) ... 739
Ourth, Lynn, & Brown, Kenneth B. Inadequate Mothering and Disturbance in the Neonatal Period ... 287

Paivio, Allan, Baldwin, Alfred L., & Berger, Seymour M. Measurement of Children's Sensitivity to Audiences ... 721
Palermo, David S. Relation between Anxiety and Two Measures of Speed in a Reaction Time Task ... 401
Peterson, Donald R., Becker, Wesley C., Shoemaker, Donald J., Luria, Zella, & Hellmer, Leo A. Child Behavior Problems and Parental Attitudes ... 151
Peterson, Donald R., Quay, Herbert C., & Tiffany, Theodore L. Personality Factors Related to Juvenile Delinquency ... 355
Pietila, Caryl. (*see* McReynolds, Paul) ... 393
Preston, Anne. (*see* Crandall, Vaughn J.) ... 261

Quay, Herbert C. (*see* Peterson, Donald R.) 355

Rapp, Don W. Childrearing Attitudes of Mothers in Germany and the United States 669
Reese, Hayne W. Level of Stimulus Pretraining and Paired-Associate Learning ... 89
Reese, Hayne W. Transposition in the Intermediate-Size Problem by Preschool Children 311
Robertson, William O. An Investigation of Maternal Concerns by Mail Survey ... 423
Rosenblith, Judy F. Imitative Color Choices in Kindergarten Children 211

Sarason, Seymour B. (*see* Barnard, James W.) 379
Sarason, Seymour B. (*see* Davidson, Kenneth S.) 199
Seidman, Stanley B. (*see* Meyer, William J.) 117
Serot, Naomi M., & Teevan, Richard C. Perception of the Parent-Child Relationship and Its Relation to Child Adjustment 373
Shoemaker, Donald J. (*see* Peterson, Donald R.) 151
Smith, Donald C., & Wing, Lucy. Developmental Changes in Preference for Goals Difficult to Attain 29
Solomon, Arthur L. Personality and Behavior Patterns of Children with Functional Defects of Articulation 731
Stevenson, Harold W., & Odom, Richard D. Effects of Pretraining on the Reinforcing Value of Visual Stimuli 739
Stevenson, Harold W., & Weir, Morton W. Developmental Changes in the Effects of Reinforcement and Nonreinforcement of a Single Response 1

Tallman, Irving. Adaptability: A Problem Solving Approach to Assessing Childrearing Practices 651
Teevan, Richard C. (*see* Serot, Naomi M.) 373
Terrell, Glenn, & Ware, Robert. Role of Delay of Reward in Speed of Size and Form Discrimination Learning in Childhood 409
Terrell, Glenn. (*see also* Ware, Robert) 789

Walder, Leopold O. (*see* Eron, Leonard D.) 457
Ware, Robert, & Terrell, Glenn. Effects of Delayed Reinforcement on Associative and Incentive Factors 789
Ware, Robert. (*see also* Terrell, Glenn) 409
Watson, Sara. (*see* Kagan, Jerome) 625
Weir, Morton W. (*see* Stevenson, Harold W.) 1
Wenar, Charles. The Reliability of Mothers' Histories 491
Werner, Emmy, & Gallistel, Elizabeth. Prediction of Outstanding Performance, Delinquency, and Emotional Disturbance from Childhood Evaluations 255
White, Sheldon H. Effects of Pretraining with Varied Stimuli on Children's Discrimination Learning 745
Wight, Byron W. (*see* Downing, Robert W.) 561
Williams, E. Jane. (*see* Kessen, William) 7
Williams, Joanna P., & Kessen, William. Effect of Hand-Mouth Contacting on Neonatal Movement 243
Williams, Joanna P. (*see also* Kessen, William) 7
Wilson, Robert C. (*see* Morrow, William R.) 501
Wing, Lucy. (*see* Smith, Donald C.) 29

Zimbardo, Philip G. (*see* Barnard, James W.) 379
Zuk, G. H., Miller, Ralph L., Bartram, John B., & Kling, Frederick. Maternal Acceptance of Retarded Children: A Questionnaire Study of Attitudes and Religious Background 525

DEVELOPMENTAL CHANGES IN THE EFFECTS OF REINFORCEMENT AND NONREINFORCEMENT OF A SINGLE RESPONSE[1]

HAROLD W. STEVENSON* *and* MORTON W. WEIR
University of Minnesota

This study investigates developmental changes in the effects of reinforcement and nonreinforcement of a response in a three-choice situation. The responses of preverbal children in learning situations (e.g., 3, 4) appear to be in line with predictions from reinforcement theories of behavior. Reinforcement of a response appears to increase the frequency with which the response is repeated and nonreinforcement to decrease this frequency. Older children and adults, however, have been found to behave in a different fashion. For example, Jarvik (2) and Gratch (1) have found evidence for a "negative recency effect" in these *S*s' responses; that is, these *S*s tend to predict that the previously occurring or reinforced event will not recur immediately and select an alternative event on a subsequent trial. There appear to be developmental changes in the effects of reinforcement and nonreinforcement on behavior.

The present investigation utilizes data collected in previously published studies (5, 6). A three choice-situation was employed in which one of the choices was reinforced and the other two choices were not. The analysis of behavior during the first few trials in this task makes it possible to determine the effects of initial reinforcement or nonreinforcement of a response on the tendency of *S*s at different development levels to repeat this response on the subsequent trial.

METHOD

Subjects

The records of 380 children were reviewed to obtain (a) 30 *S*s at each age level who were not reinforced on their initial trial and (b) 30 *S*s at each age level who received reinforcement at the latest on their fourth choice

---

* Institute of Child Development and Welfare, University of Minnesota, Minneapolis 14.

[1] The study was supported by a grant from the National Science Foundation (Grant 3280) to the senior author.

Child Develpm., 1961, 32, 1-5.

of the reinforcing stimulus. Children at four age levels were selected: 3-0 to 3-11, 5-0 to 5-11, 7-0 to 7-11, and 9-0 to 9-11 years.

Apparatus

The apparatus has been described previously (6). It consisted of a yellow panel containing a horizontal row of three knobs. Above the knobs was a signal light and below the knobs was a delivery hole for marbles. The marbles fell into an enclosed plastic container.

Procedure

An identical procedure was followed in the studies from which the data for this investigation were collected. Essentially, it consisted of *S*'s being instructed to push one of the three knobs when the signal light came on and being told that if he pushed the correct knob a marble would fall into the plastic container. One of the three knobs was designated the correct knob for a particular *S* and, depending upon the condition under which *S* was tested, either 33, 66, or 100 per cent of the choices of this knob were reinforced. The other two knobs never yielded reinforcement.

RESULTS

The record of each *S* was examined to determine the response of *S* on the subsequent trial when (a) *S*'s initial response in the task was not reinforced, (b) *S*'s initial response to the reinforcing stimulus was reinforced, and (c) *S*'s response to the reinforcing stimulus was reinforced for the first time. The latter two conditions differ in that in (b) *S*'s first response to the reinforcing stimulus was reinforced, while in (c) *S*'s initial response to the reinforcing stimulus was not reinforced but a later response was reinforced. No more than three nonreinforced responses to the reinforcing stimulus occurred in the latter case prior to the first reinforced trial. The subsequent response was classified according to whether or not the prior response was repeated by *S*.

Effects of Nonreinforcement

Performance on trial 2 was examined for *S*s who were not reinforced for their response on trial 1. The proportion of *S*s at each CA level who changed response on trial 2 is presented in Figure 1. The *S*s at CA 3 show some tendency to repeat the nonreinforced response, while very few *S*s at the older age levels show this tendency. The difference between the performance of *S*s at CA 3 and the other groups combined is highly significant ($\chi^2 = 24.87$, $df = 1$, $p < .001$).

Effects of Reinforcement

It is possible that performance on the subsequent trial might differ for *S*s who were reinforced on their initial choice of the reinforcing stimulus and for *S*s who were not reinforced for their choices of the reinforcing

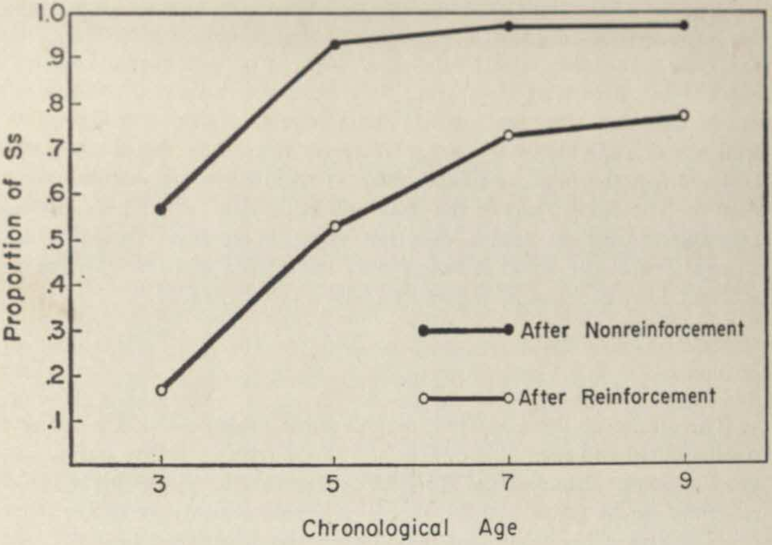

FIGURE 1—Proportion of *S*s at each CA level changing response following reinforcement and nonreinforcement of the prior response.

stimulus on several trials prior to receiving their first reinforcement. In order to determine whether these two conditions differ, a comparison was made of the *S*s in situations (b) and (c) described above. Of 40 *S*s (10 at each CA level) who were reinforced on their first choice of the reinforcing knob, 22 repeated the response on the next trial, while of 80 *S*s (20 at each CA level) who received reinforcement only after choosing the reinforcing stimulus up to three times with no reinforcement, 32 repeated the response on the next trial. The difference between these two conditions is not significant ($\chi^2 = 1.86$, $df = 1$, $p > .10$), and for further analysis the *S*s in the two situations were combined.

The proportion of *S*s at each CA level who changed response following reinforcement is presented in Figure 1. A consistent increase through the four age levels is seen in the number of *S*s changing response. The increase with age is more gradual than that found in the analysis of the effects of nonreinforcement. The results for the older *S*s provide clear evidence for the operation of a negative recency effect. The differences among the four age groups are highly reliable ($\chi^2 = 27.61$, $df = 3$, $p < .001$), indicating significant changes in the effects of reinforcement as a function of CA.

Data for Adult Subjects

It is of interest to compare the performance of the older children with that of adults (college students), for whom data were available in the same situation under the same experimental conditions. The number of *S*s changing response following their first reinforced trial continued to increase from the 9-year to the adult level. Of the 30 adult *S*s, 90 per cent changed response on the subsequent trial. Comparing these results with those in Figure 1, it can be seen that after the age of 7 there is relatively little change in the tendency of individuals to change response after their initial reinforced trial and that the negative recency effect is manifest in the performance of most *S*s. The comparison of the child and adult data for trial 2 following nonreinforcement on trial 1 does not show a consistent developmental change. Ten of the 30 adult *S*s and only one of the 30 9-year-old children repeated the nonreinforced response on trial 2 ($\chi^2 = 7.13$, $df = 1$, $p < .01$).

DISCUSSION

The results of these analyses provide clear evidence that the effect of reinforcement and nonreinforcement of a single response differs significantly as a function of chronological age. The behavior of the 3-year-olds is readily accounted for by traditional S-R concepts. Reinforcement increased the tendency to repeat the reinforced response on the subsequent trial and nonreinforcement reduced this tendency. By CA 7, reinforcement of a response did not increase the tendency to repeat the response; in fact, the effect appeared to be just the opposite, for *S*s tended to change their response following reinforcement to nearly as great a degree as they did following nonreinforcement.

These results do not, of course, indicate that an S-R analysis of the behavior of older children is impossible. They do indicate that some other unit of response must be utilized in discussing older *S*s' behavior. It may be assumed that the behavior of older children is guided by verbal hypotheses and that it is in the confirmation of these hyoptheses that reinforcement has its greatest effect. Children's verbalizations of their basis of response seem to indicate that one of the dominant hypotheses in a multiple-response situation is that there is a pattern of responses which leads to reinforcement. An analysis of behavior in terms of patterns or sequences of responses may indicate that nonreinforcement and reinforcement have the same effect on these units of response for older *S*s as they do on the single responses of younger *S*s.

SUMMARY

Data for 30 *S*s at each of four age levels from 3 through 9 years were selected to determine the effects of initial nonreinforcement and reinforcement of a single response on behavior during the subsequent trial. Significant differences were found in the effects of both reinforcement and nonreinforcement as a function of CA. A strong tendency was found for the negative recency effect to increase with increasing CA. The results were

interpreted as indicating that utilization of a single S-R unit is inappropriate in the analysis of the behavior of older children and adults.

REFERENCES

1. GRATCH, G. The development of the expectation of the nonindependence of random events in children. *Child Develpm.,* 1959, 30, 217-228.
2. JARVIK, M. H. Probability learning and a negative recency effect in the serial anticipation of alternating symbols. *J. exp. Psychol.,* 1951, 41, 291-297.
3. KUENNE, M. R. Experimental investigation of the relation of language to transposition behavior in young children. *J. exp. Psychol.,* 1946, 36, 471-490.
4. SIEGEL, P. S., & FOSHEE, J. G. The law of primary reinforcement in children. *J. exp. Psychol.,* 1953, 45, 12-15.
5. STEVENSON, H. W., & WEIR, M. W. Variables affecting children's performance in a probability learning task. *J. exp. Psychol.,* 1959, 57, 403-412.
6. STEVENSON, H. W., & ZIGLER, E. F. Probability learning in children. *J. exp. Psychol.,* 1958, 56, 185-192.

SELECTION AND TEST OF RESPONSE MEASURES IN THE STUDY OF THE HUMAN NEWBORN[1,2]

WILLIAM KESSEN,* E. JANE WILLIAMS, *and* JOANNA P. WILLIAMS
Yale University

Thirty years ago, Weiss stated the problems of response selection in studies of the human newborn:

> . . . There is very little coordination of the different movements. . . . The assumption that no movement is ever *repeated exactly* becomes a very obvious fact. It seems at first futile to speak of classifying the behavior. While this is true theoretically, as a matter of fact we do group together certain types of movements according to their resemblances (29, p. 454).

Unfortunately, Weiss's criterion of classification—the occurrence of "resemblances"—does not provide an unambiguous rule for ordering the behavior of newborns. Should we attend to resemblances in anatomical topography, or speed, or temporal overlap, or effect on the environment, or relation to antecedent stimulation? In fact, all of these criteria for classification have been used by students of neonatal behavior, with little evidence that a simple general rule for response definition is emerging. The problem of response definition is by no means confined to research with infants—it has long been an issue in general psychology (19, 22)—but the newborn presents special difficulties in answering this most basic of research questions: What are the effective response-classes?

In the observation of animals and older human beings, psychologists of all theoretical commitments depend on *movement in relation to the environment* and on *speech* to provide the most elemental definition of response categories. Whether in the study of maze-learning or of psychotherapy, the

---

* Department of Psychology and Child Study Center, Yale University, 333 Cedar Street, New Haven, Connecticut.

[1] The research reported here was supported in part by a grant from Fluid Research Fund, Yale University School of Medicine, and in part by Research Grant M-1787, United States Public Health Service.

[2] The writers are deeply indebted to Dr. Milton J. E. Senn and to the staff of Grace-New Haven Community Hospital for their generous cooperation. S. J. Onesti, J. Doris, J. C. Beck, A. L. S. Hendry, and A. M. Leutzendorff assisted in the collection and analysis of data during the research and in observations preliminary to it.

Child Develpm., 1961, 32, 7-24.

psychologist of mature behavior can make use of the fact that his subjects move through space or speak; the student of newborns can claim no such advantage. This difference turns out to be quite important in the first stages of the search for meaningful response categories. Although general definitions of response-classes have been developed (e.g., those responses which increase in probability of occurrence with the occurrence of reinforcement), a great deal of psychological observation still depends on the use of categories common to "natural" everyday usage. Thus, the behavior of picking up a cigaret and the behavior of picking up a rattle can "naturally" be put in the response-class Picking Up, even though the responses may differ along many dimensions. Similarly, the expression "I hate dishwashing" and the expression "Golf is repulsive to me" will naturally indicate the same response-class of Disliked Activities even though they are phonetically quite different and may differ on other grounds as well. People organized the activities of other people into response-clusters long before there were professional psychologists, and these vernacular rules for the classification of behavior seem to depend in large measure on characteristics of language and on regularities in response to objects (including other organisms) in the environment. When the psychologist sets out on his systematic study, he almost invariably adapts his observation to these everyday classificatory rules. Such a strategy is reasonably successful when used with mature human beings and even with animals, but in most cases of newborn behavior the "natural" rules are inapplicable. It is largely for this reason that the behavior of newborns is seen as meaningless; we are unable to organize the neonate's behavior with the rules of classification natural for older organisms.[3]

Curiously enough, the lack of obvious criteria of meaningfulness in newborn behavior also precludes exclusion of response-classes; that is, any behavior might be productively studied. This ambiguity, which is a problem for all psychologists but of special importance for those of us interested in early behavior, forces attention to the more general question: What do you want the response-categories for? There are several ways in which this question has been answered in the past, and a brief review of some of the positions taken will put the issue of response selection in a sharper light.

In the tradition of comparative psychology, the human newborn can be seen as a separate species, or at least a separate breed, of interest to the psychologist because of his variation from the behavior of other animals. Much of the classical research on neonatal behavior fits with this view of the infant. Can he discriminate color (27)? What are his taste sensitivities (17)? Can he learn (24)? Is he responsive to privation (16)? What reflexive behavior does he show (2)? Questions of this order, which led to the accumulation of important observations of infant behavior, were almost invariably derived from general psychology; those independent variables were manipulated which had been important in the study of other animals, and response definitions were developed which showed more-or-less reliable relationships with external stimulation. Dennis's (2) survey of the response repertory of the human infant lists the response-classes which resulted from this pioneer work on neonatal reflexology, and his summary is convincing demonstration that a relatively large number of response categories for neonatal behavior can be described under the criterion of close temporal relation to external stimulation.

---

[3] Wolf (30) has suggested that the basic natural rule for organizing the behavior of human beings depends on the observation of "goal" or "endpoint" behavior. If this is the case, then the apparent meaninglessness of the newborn's activities may represent the inability of adult observers to recognize the newborn's "goals."

In the course of early research on newborn behavior, there was developed a response-class very different from the simple "reflexive" activities which dominate Dennis's list. Called variously "mass activity" or "general activity," this response category was the antithesis of the stimulus-response descriptions and had as its criterion of inclusion *any* detectable movement of the newborn. Several techniques were used to determine measures for this response-class, but in general they aimed at representing all movements of the infant, whether these movements, collected in smaller subclasses, could be reliably related to external stimulation or not. For example, a response class defined in this way was the chief tool in Irwin's fundamental work describing the effect of several antecedent manipulations on neonatal activity (e.g., 15). There was promise, during the thirties, that "general activity" would serve for the study of newborn behavior the same broad function that "pressing a bar" or "running a maze" had served for other psychological problems. Perhaps because the specific measures devised for neonatal activity were not sensitive enough to detect response changes of brief duration, this promise was not fulfilled.

The study of the newborn as species, directed more by interest in reliable description than by commitment to elaborated theoretical issues, produced a natural history of neonatal behavior which was fundamental to later study and developed a response class—general activity—which was shown to be responsive to manipulation.

Running alongside research interest in newborns as behavioral variants, and never independent of it, has been the study of neonatal behavior as the precursor of later behavior. In this view, the newborn is seen as part of a long development toward adult human functioning, and response-classes have been sought which will show continuity of this development over time. However, as one might expect from our earlier discussion, there are no obvious and natural continuities between the behavior of newborns and the behavior of older human beings. As a result, the search for genotypic continuity in the face of phenotypic diversity (30) has led to a stronger emphasis on theory—that is, on postulated continuity—than was the case for the comparative psychologists of infancy. From this search for developmental laws have come the three most important contributions to a general theory of newborn behavior—Freud's theory of personality development (e.g., 3), Gesell's theory of maturation (e.g., 9), and Piaget's theory of the growth of intelligence (e.g., 26). Each of these systems, devised to deal with the entire range of human development, makes statements about the important response-classes to be observed in the study of newborns.

Gesell's work has had by far the widest research impact, leading to a line of studies on the normative development of the human being, particularly with regard to relatively simple motor behavior (e.g., 10, 13). Piaget's interest has been in aspects of neonatal behavior which show the operation of what he considers general intellective functions, and his observations (26) have as a theme the infant's development of increasingly effective ways of dealing with the environment. To date, no systematic study has been made of newborn behavior with Piaget's postulations as guide. The developmental theory outlined by Freud, which emphasized the early expression and control of impulse, has produced a great deal of commentary (4, 12, 14) and some provocative research (e.g., 28). Its implications for the study of neonatal behavior have been treated in detail by Lustman (23), who also discusses the data on newborn behavior relevant to psychoanalytic theory.

Whatever theoretical position is taken, the study of the immature organism as precursor of the mature organism involves two interrelated problems: the establishment of developmental sequences common to all human beings (e.g., changes in prone progression, psychosexual development) and the determination of parameters which permit the description of individual differences (e.g., accelerated development, regression rate). In the specific instance of the newborn, one may ask: What behavior or change in behavior is common to all infants (e.g., occurrence of sucking movements when orally stimulated). Or one may ask: What behavior or change in behavior is characteristic of one child or of a limited group of children (e.g., variations in intensity of sucking movements). Although the first kind of normative question receives continuing examination, research interest in the study of newborns has shifted in recent years to the second. The investigations of Fries (5, 6) on activity, of Balint (1) on sucking style, of Lustman (23) on response to tactile stimulation, and of Graham *et al.* (11) on sensitivity, among others, reflect an increasing interest in the dimensions of variations among human newborns.

In short, the consideration of the newborn as a precursor brings to the problem of response selection two points of emphasis: the relevance of neonatal behavior to later behavior and the detection of individual differences. Attention to these aspects of the problem also makes clear that the lack of obvious and common-sense continuities between neonatal and later behavior requires that the selection of response-classes be directed in the first instance by hunch or preconception or theory.

Once a response-class has been selected for study, however, whether on the basis of a specific developmental theory or on the basis of a less well-organized hunch, criteria of empirical relevance become important, and the usefulness of a response-class in the systematic study of newborn behavior comes to be determined by factual considerations. Can the response be observed with high intersubjective agreement? If it is held to indicate individual differences, are values of the response stable for a neonate from occasion to occasion? Is the response related in a systematic way to age, antecedent conditions, or other responses?

The problem of classification posed by Weiss goes beyond the statement of "resemblances," both because the resemblances are not always obvious and because the resemblance selected may not meet empirical requirements for a meaningful response category. The selection of response-classes in the study of newborn behavior will be at first directed by hunch or theory, and those response-classes which are shown to be reliable, stable, and systematically related to other events will survive as useful parts of a system of newborn behavior.

The research to be reported in part here grew from an interest in the early history of tension and tension-control. As Wolf has pointed out, the history of the child can be seen as the history of his responses to developmental "problems," and "the earliest behavioral manifestations of problem-solving are the baby's modes of tension reduction" (8, p. 507). A closely related concern with tension or disequilibrium as important in the understanding of behavior is reflected in the notions of "drive" in contemporary learning theory (e.g., 25) and of "instinct" in psychoanalytic theory (4). In the apparent chaos of neonatal behavior, two primitive generalizations related to these theoretical notions seem to be warranted. First, there are changes from occasion to occasion in the tension or disturbance or disequilibrium of the human newborn. He may be quietly asleep and the next moment awake and screaming; he may be slowly moving his arms and legs and then suddenly startle; he may produce sighs of contentment in the observer or twitches of distress. The second generalization about which we can be relatively certain is that these early variations in tension come to be more and more closely related to environmental events. In fact, to the degree that one is able to set up a stable relation between some aspect of the environment and changes in the infant's level of tension, to that degree will his behavior become meaningful. The infant may become quiet when lifted to the feeding position, for example, or begin to cry "for food" at regular temporal intervals. These basic observations, coupled with the theoretical importance of tracing the early history of tension, led to the first guiding question of the present research: What variables or events influence variation in neonatal tension and tension-control, and what is the early developmental course of this influence?[4]

---

[4] There are two other closely related ways in which early changes in human behavior can be discussed. The child may be seen as increasingly *in control of his environment,* that is, as becoming more competent. Alternatively, the child may be seen as coming to be increasingly *under the control of his environment,* that is, as becoming more predictable in relation to external events. The differential implications of these two statements of the problem and the "tension" statement will be dealt with in a later paper.

Beyond variations in tension and tension-control in the same children from occasion to occasion, there are theoretical (7, 8) and suggestive empirical (30, 31) grounds for believing that reliable variation in tension occurs between subjects, that is, that tension is a source or index of individual difference. Moreover, in the case of developing tension-control, there is reason to believe that different infants may show different tension-control patterns, some children being more responsive to tactile stimulation, others to gastrointestinal changes, and so on (21). From these considerations emerged the second general question of the research: Can stable individual differences in patterns of tension and tension-control be described for the newborn?

Clearly, these most general statements of the problem do not provide obvious and natural response measures; there remains a wide range of observational protocols which might serve as indicators of tension, and there are perhaps even more variables to be investigated as important in tension-control. For the first empirical analysis of the problem, therefore, the study was narrowed to an investigation of general movement and crying as indicators of tension and of mouthing and hand-mouth contacting as potential tension-control activities. The choice of crying was based on the almost universal popular generalization that it is an expression of distress by the infant; it most assuredly distresses the adult observer. The use of general movement as a partial definition of tension is also justified in common-sense psychology—"a quiet baby is a happy baby"—but it receives support from other sources as well. General movement or activity has been ascribed substantial theoretical importance in an understanding of neonatal functioning; as mentioned earlier, measures of movement have had empirical use of some success; and perhaps of greatest relevance to long-term research goals, we hoped to develop a measure of movement which would be sensitive to changes of short duration in the behavior of the newborn.

The selection of oral behaviors as possibly relevant to tension-control was dictated by both observation and theory. Whatever other disagreements may separate theorists of early development, all agree that mouthing and sucking are fundamentally important responses in the newborn. Moreover, preliminary observation had suggested that two oral responses—hand-mouth contacting and mouthing—could be reliably observed and might reflect systematic individual differences.

Thus, the larger questions raised earlier become reduced to the more readily testable problem: What events or variables control variation in crying, general movement, mouthing, and hand-mouth contacting in the neonate, and how does this control change during the first days of life? There are many routes along which this question can lead, but our initial step was to safeguard more complicated issues by first determining whether the measures chosen meet the empirical criteria stated earlier, that is, by assessing the reliability, stability, and relatedness of our measures. Such is the aim of the present report. Can general movement, crying, and aspects of oral behavior be measured reliably and sensitively? Do these behaviors show consistent individual differences? Do these behaviors show systematic relations with each other and with situational variables?

METHOD

Subjects

The subjects of the research were mature, normal infants in the maternity service of Grace - New Haven Community Hospital. Each subject was seen several times during his hospital stay; most subjects were observed at least once each day (*see* below, *Schedule of Observations*). In order to eliminate systematic bias in the selection of subjects, the following criterion was established for inclusion in the sample: that, whenever we began an observation series, the subject chosen be the youngest newborn in the nursery who had not yet been fed. Thus, the age of subjects at the time of their first observation ranged from 1½ hours to 13½ hours, with the median age at 8½ hours. This criterion of selection had the further advantage of permitting the comparison of neonatal behavior before and after the first feeding.

The present report is based on the repeated observation of 50 newborns. The specific observational procedures used were changed as the research became more focused and detailed, with the subjects falling into two main groups: group I ($N = 34$), where our chief interest centered on the observation of hand-mouth contacting; and group II ($N = 16$), where a detailed examination was made of mouthing, crying, contacting, and movement. Each of these groups was subdivided to provide information on special problems (*see* below, *Schedule of Observations*).[5]

Schedule of Observations

Typically, each infant was observed on each of the five days of his hospital stay. Shortly before noon, the infant was brought from nursery to the observation room, which was an unoccupied nursery on the same floor of the hospital. Blankets and other wrappings were removed, and the newborn (dressed in hospital shirt and diaper) was placed supine on the observation pad. If the child was crying, he was held or rocked in arms until he quieted. Once the observation was begun, however, the child was not handled except in cases of acute distress, when a nurse was called to tend to him. The observation period covered from 10 to 30 minutes, depending on the group to which the child was assigned, after which the child was returned to nursery. The schedule of observations for each group of subjects was as follows:

> Group Ia ($N = 25$): General observation, followed by five-minute observation (checklist) by at least two observers of hand-mouth contacting. This procedure was repeated, when possible, on days 1, 2, 3, 4, and 5. For 10 sub-

---

[5] The limitations of the clinical setting in which the research was conducted and occasional apparatus failures led to the discarding of some data on some subjects. As a result, in the analyses presented later, *N*s will vary from those given here.

jects in this group, a pilot run of motion photography (30 seconds in duration) was made on three successive days.

Group Ib ($N = 9$): General observation, followed by five-minute observation (checklist) by at least two observers of hand-mouth contacting. This procedure was repeated, when possible, three times each day (at 8-9 a.m., at 11-12 a.m., and at 5-6 p.m.) on days 1, 2, 3, 4, and 5.

Group IIa ($N = 11$): General observation, followed by five-minute observation (key-depression with record made on an Esterline-Angus operations recorder) of crying, mouthing, and hand-mouth contacting. After this observation, the same measures were taken over five one-minute observation periods, each separated from the next by one minute of no recorded observations. During each of the one-minute observation periods, a 30-second sample of motion photography was obtained. For 10 to 15 seconds of minute-observations 8 and 12, the observer inserted his shielded index finger into the baby's mouth. Figure 1 shows this procedure in diagrammatic form. The procedure was repeated on days 1, 2, 3, 4, and 5.

Group IIb ($N = 5$): Procedure was exactly the same as that for group IIa, except that no oral stimulation was applied and observations were not made during minutes 8, 10, 12.

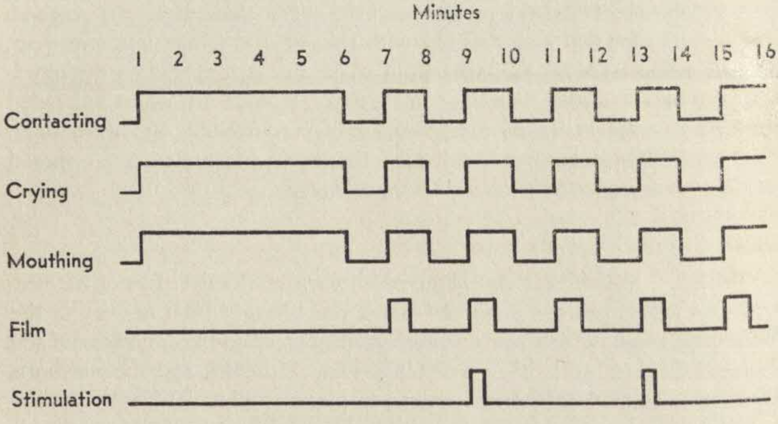

FIGURE 1—Schedule of observational procedures, group IIa.

Response Measures

Hand-mouth contacting. The newborn's response of bringing his fingers or hands to his mouth was recorded in two ways. The first, used with group I, involved placing a checkmark on a mimeographed sheet whenever a contact occurred, recording thereby the finger which made contact and the *frequency* of contacts over a predetermined period of time. For group II, the depression of one of two telegraph keys marked a line on an Esterline-Angus operations recorder (Recorder), indicating which hand made contact, the *frequency* of contacts over a predetermined period, and the *duration* of hand-mouth contacting.

Mouthing. In addition to the movements of the newborn's mouth which can be observed in the feeding situation and which frequently accompany hand-mouth contacting, mouthing movements occur in the newborn in the absence of external stimulation. For group II, a telegraph key was depressed, marking a line on the Recorder whenever the infant's lips moved in a characteristic sucking pattern, either in the presence or absence of external stimulation. In the analysis presented below, the *duration* of mouthing is the measure of this activity.

Crying. For group II, a telegraph key was depressed, marking a line on the Recorder whenever the infant made an audible cry. In order to avoid observer anticipation of crying, the recording key was depressed only on the exhalation part of the crying cycle. As in the case of mouthing, *duration* of crying is the measure used for analysis.

For all duration measures, the Recorder tape was scored by seconds of observation, and a score of 1 was assigned for each second in which hand-mouth contacting, mouthing, or crying was recorded. Thus, for all three duration measures, an infant's score could vary from zero to whatever number of seconds elapsed during the observation. For comparability among observations of different lengths, all scores were reduced to a "per 30-seconds" base. Therefore, for any observation, an infant's crying, mouthing, or hand-mouth contacting duration score varied in a range from zero to 30.

General movement. Each child in group II and 10 children in group I were photographed by a fixed-position overhead camera for one or more runs of 30-seconds duration on several days of the lying-in period. The whir of the camera's spring drive did not, by and large, alter the infant's behavior. No artificial illumination was used other than the overhead lights permanently installed in the nursery. These filmstrips were analyzed to provide a measure of the horizontal displacement of hands and feet over five-second intervals. The details of the technique are described elsewhere (20); although the technique is tedious and time-consuming (requiring the examination and measured comparison of 50 film-frames per child per observation), it provides a reproducible and permanent record of the infant's movements.

RESULTS

Reliability of Measures

Hand-mouth contacting. The reliability of the checklist technique was high; correlations between pairs of observers were invariably above .90, and, for the two observers who did the bulk of observations for group Ia, the interobserver correlation was .96. The Recorder technique used for group II showed similar intersubjective reliability, with the protocols of the two most frequent observers showing a correlation of .96.

Mouthing. The occurrence of mouthing was the event most difficult to discriminate in the study. Nonetheless, test runs for reliability produced an interobserver correlation coefficient of .93.

Crying. Although there is some ambiguity in the observation of crying (e.g., unvoiced cries, "crying face" without sound), observers did not have the reservations typical for the observation of mouthing, and test runs for reliability showed an interobserver correlation of .99.

Movement. The assessment of reliability for the movement measure was made by comparing the movement values assigned to children by two analyzers after they had independently sketched and measured the appropriate frames of film. The discriminations required for this task were matching one picture against the next and measuring linear displacement of limbs with a millimeter rule [*see* (20) for further details]. The correlations which resulted from this assessment ran above .96.

The measures set up for general movement, crying, mouthing, and hand-mouth contacting were shown to be very highly reliable; moreover, the absolute values recorded by different observers did not show more than negligible variation in means and variances.

Individual Stability of Measures

Being assured that reliable protocols can be established for movement, crying, hand-mouth contacting, and mouthing, we put this next question to the data: Are these measures stable from infant to infant over the lying-in period; that is, is there evidence of consistent individual variation among newborns for these response categories?

Hand-mouth contacting. Hand-mouth contacting is by no means an infrequent response in the human newborn. Without exception, every infant seen in pretest series and in the research proper was observed to bring his hand to his mouth at least once during the five-minute observation periods. Moreover, the variation in frequency of the response is great; among the 32 subjects who were observed on each of the five days of hospital stay, the average number of hand-mouth contacts during five minutes ranged from 3 to above 30, with two subjects making contact 60 times in a single observation period.

In group I (checklist group), 16 newborns were seen on all five days at the midday observation period; this provides an array of 80 contacting scores which can be evaluated in two ways to indicate individual stability. The first of these is a rank order procedure, in which frequency of hand-mouth contacting is ranked for each day, and the coefficient of concordance (18) computed for the resulting array of ranks. This statistic had a value of .51 ($p < .01$), indicating that a significant proportion of variation in ranked frequency of hand-mouth contacting could be assigned to individual differences. The same finding can be shown in an analysis of variance of the day-by-subject array of hand-mouth contacts; such an analysis shows a significant variance ($F = 3.47$, $df = 15/60$, $p < .01$) ascribable to between subject variation.

Group II (Recorder) provided an opportunity to check the replicability of this finding. Sixteen subjects were seen on five days of hospital stay, and the resulting coefficient of concordance was once again significantly different from chance ($W = .34$, $p < .05$). The analysis of variance of raw contacting scores showed a parallel outcome ($F = 2.19$, $df = 15/60$, $p < .05$).

In evaluating these findings, it must be remembered that they are based on a relatively brief sample of the infant's total stay in the hospital (about 0.5 per cent), and under circumstances (e.g., variation in maternal handling) which introduce considerable unsystematic variation. Therefore, the stability demonstrated here is almost certainly an underestimate of the stability to be expected from a longer time sample or in an artificially controlled setting.

It is interesting to note that the data from group II provided no evidence of significant individual stability for *total duration* of hand-mouth contacting ($W = .22$) nor for *mean duration per contact* ($W = .10$). This result suggests that getting the hand to the mouth may be under the control of variables different from those which control the maintenance of a contact once made.

Mouthing. Some activity of the mouth, whether responsive to external stimulation or not, is a strongly prepotent response in the human newborn. Among the 16 infants in group II, the five-day average duration of mouthing during a five-minute observation ranged from 56.4 seconds (18.8 per cent of time observed) to 189.6 seconds (63.2 per cent) with a mean at 112.5 seconds (37.5 per cent). For this measure, a coefficient of concordance computed on an array of ranks does not reach statistical significance ($W = .21$); the corresponding analysis of variance shows a marginally significant variance attributable to between subject variation ($F = 2.09$, $df = 15/60$, $p < .05$). For the conditions of observations obtaining in the present study, mouthing is apparently not as sensitive a measure of stable individual variation as is hand-mouth contacting.

Crying. For the 16 subjects in group II, the crying measure showed wide variation, with mean duration per five-minute observation over the lying-in period ranging from 6.8 seconds (2.3 per cent of time observed) to 117.6 seconds (39.2 per cent) with a mean at 54.5 seconds (18.2 per cent). The distribution of crying durations was irregular in shape, with 27 of 80 observations showing no crying at all. No significant individual differences were found on the crying measure. General observation of the infants suggested that this response is both too infrequent and too much under the control of fluctuating external stimulation (e.g., wet diapers) to be a convincing index of individual variation.

Movement. The 30-second filmstrips run for each subject in group II on each day of his hospital stay provided a measure of neonatal movement. The sum of displacements over five-second intervals of right and left palm and right and left ankle was the primary score submitted to analysis. For 10 subjects in group IIa, this score was computed for five daily observations and these scores combined to provide a single daily value of movement for each infant. The resulting coefficient of concordance was .53 ($p < .01$), confirming the general observation that a significant portion of the variation in newborn movement can be assigned to individual variation. The corresponding analysis of variance shows the same result ($F = 3.86$, $df = 9/36$, $p < .01$). It should be noted, too, that this score combines observations of "spontaneous" movement with observations of movement following external stimulation (i.e., insertion of the observer's shielded finger in the infant's mouth at observation minutes 8 and 12 each day). Individual stabilities were clearly potent enough to override the effect of this intervention, and a high degree of confidence can be placed on the finding of consistent subject differences.

This stability in neonatal movement can be seen even more dramatically by examining the scores of groups IIa and IIb for the *first* 30-second observation each day. The coefficient of concordance for 15 subjects was .49 ($p < .01$); the analysis of variance showed significant variation between subjects ($F = 3.33$, $df = 14/56$, $p < .01$). Thus, a stable measure of individual variation in newborn movement can be computed, *based on a 30-second sample* on each day of hospital stay.

The range of variation among neonates in movement over a five-second measured interval, averaged for the lying-in period (groups IIa and IIb, first filmed observation each day) was from 1.50 cm./limb to 5.70 cm./limb, with a mean at 3.65 cm./limb.

In assessing the four measures under examination in the present study for their sensitivity in showing stable individual variation, we may conclude that hand-mouth contacting and general movement vary in a significant way from child to child, that mouthing is a marginally sensitive measure of individual variation, and that crying as measured in the present setting does not provide a reliable index of individual variation.

Relatedness of Measures

Day-to-day effects. During the first five postnatal days, the newborn does not merely age; he is trained to a new feeding technique, he is extensively handled, if he is a male he undergoes minor surgery, and so on. It is therefore difficult to assign sources of variation in day-to-day effects. However, several meaningful questions may be asked. Does the behavior show regular and systematic change over days? Are there significant differences between day 1 (when, for our sample, no feeding had occurred) and later days? Do relationships among variables change over time? There are 15 subjects in group II on whom complete data are available which can be used to answer these questions.

Figure 2 shows changes over days in mouthing, hand-mouth contacting, and crying. As might be anticipated from the earlier finding of great variability of mouthing and crying, neither of these measures shows significant day effects. Whether the increase in crying scores between day 1 and day 4 which appears to visual inspection can be brought to significance by a larger sample, whether of subjects or time, remains to be investigated. Of more interest is the finding that mouthing is insignificantly different from day 1 to later days. It might have been anticipated that feeding experience would affect mouthing, either increasing it as a function of the reinforcement operation of feeding or decreasing it as a function of the increasing frustration of nonnutritive mouthing. Neither effect is evident.

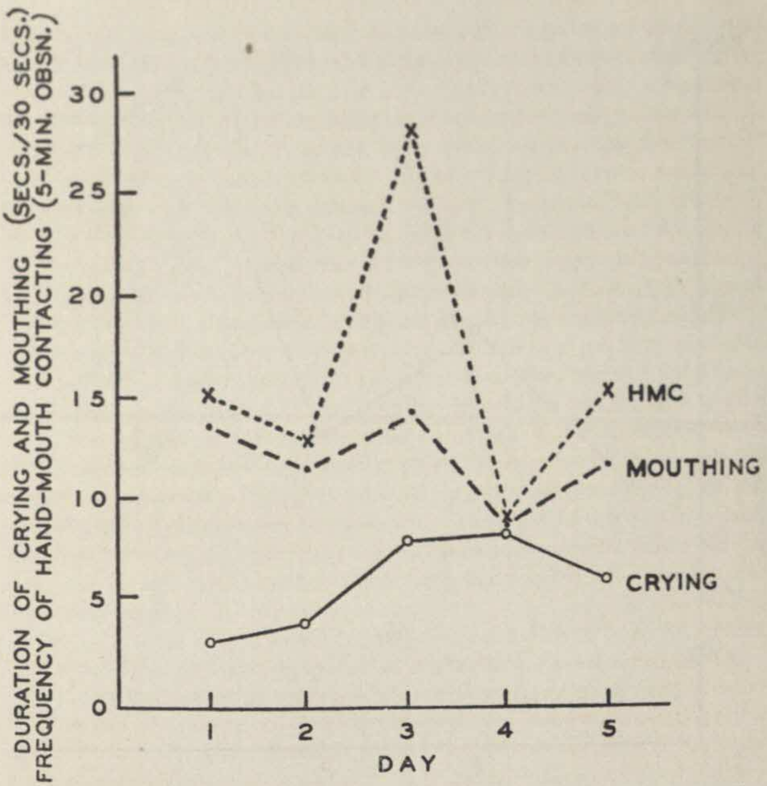

FIGURE 2—Changes in hand-mouth contacting, mouthing, and crying over the first five days of life.

The hand-mouth contacting measure is somewhat more difficult to interpret. The differences shown in Figure 2 are significant ($F = 4.51$, $df = 14/56$, $p < .01$), although it is apparent that the sharp increase in contacting on day 3 accounts for most of the variation. An examination of other responses and of hospital routines has shown no covariation which would readily account for this surprising hump. Nonetheless, it is interesting that, as in the case of mouthing, the increasing number of feeding "trials" the child has does not seem to influence hand-mouth contacting in any systematic way. Contacts on day 1 are as frequent as those on day 2, day 4, or day 5. This finding is important in that it casts serious doubt on an assumption that hand-mouth contacting is "secondary" to feeding or learned in the feeding situation.

The day-to-day plot of movement for group IIa subjects shown in Figure 3 is relatively free of ambiguity. There is a systematic and significant ($F = 3.03$, $df = 4/36$, $p < .05$) increase in neonatal movement during the early days of postnatal life. This finding, beyond confirming the observations of other investigators, adds to our confidence that the movement measure used in the present research is a sound measure of individual and situational variation.

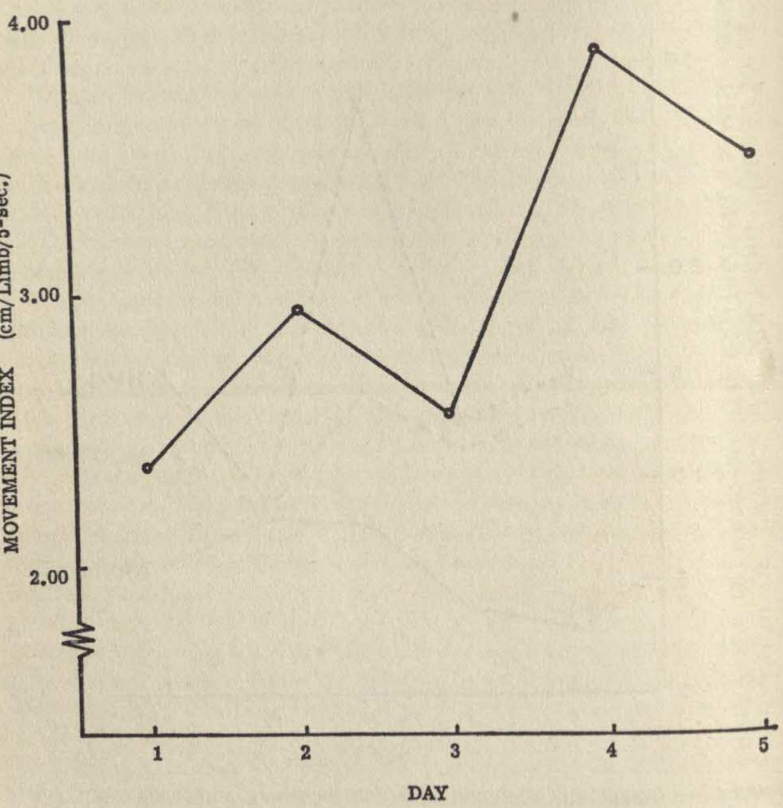

FIGURE 3—Changes in neonatal movement over the first five days of life.

Interrelations among measures. Do the response measures hang together in a way which would suggest a primitive "typology" of newborns? For each child in group II ($N = 15$), a single index was computed of crying, mouthing, contacting, and movement by summing daily values over the entire hospital stay. The rank correlations (rho) among these *over-all* measures were as follows:

Contacting-Mouthing	.36	Mouthing-Movement	.49
Contacting-Crying	.19	Movement-Crying	.18
Contacting-Movement	.19	Mouthing-Crying	—.27

Although none of these correlations reaches conventional levels of statistical significance, there is a suggestion that the more active infants mouth more and that the infants who mouth more often make contact more often. There is, furthermore, no strong evidence that a simple "general factor" will account for the relations suggested. It is interesting to note that the low correlation between movement and contacting argues against a random-hit interpretation of hand-mouth exchange; it does not seem to be the case that an infant who moves a lot happens to strike his mouth frequently.

The complexity of interrelations among newborn responses is seen in an examination of day-to-day changes in correlations (Table 1). For the comparisons (again, the coefficient is rho) crying was not included because of the relatively high incidence of zero scores. The major implication of this table is clear; a tendency for the three measures to be related fairly closely early in the postnatal period dissipates toward chance by the fifth day of life. Two hypotheses can be advanced to account for this striking shift. The interrelation at birth and shortly thereafter may be under the control of some transiently effective variable (e.g., maternal anesthetic), or the environmental contingencies of the first five days may be such as to disturb a congenitally organized response-set. The decay in interrelations among measures cannot be assigned to increasing unreliability of measures. For both movement and frequency of contacting, the *individual* stabilities between day 4 and day 5 are as great as or greater than the stability among earlier days. An accurate interpretation of these findings must await further empirical examination of newborn behavior and its environment.

TABLE 1

RANK-ORDER CORRELATION COEFFICIENTS (*RHO*) FROM DAY TO DAY AMONG THREE MEASURES OF NEWBORN BEHAVIOR

	Movement-Contacting	*Movement-Mouthing*	*Contacting-Mouthing*
Day 1	.57*	.55*	.68**
Day 2	.25	.35	.59*
Day 3	—.09	—.21	.70**
Day 4	—.16	.17	.32
Day 5	.11	—.03	.18

NOTE.—One asterisk indicates significance at .05 level, two asterisks significance at .01 level. $N = 16$ for days 1, 2, 3, and 4; $N = 13$ for day 5.

LIBRARY
1-9-70
9247

The Effects of Stimulation

On two observations each day, the infants in group IIa were given an opportunity to suck on the observer's shielded index finger for 10 to 15 seconds. Under the conditions of the present study, no stable immediate effect on neonatal movement could be detected; although the infants seemed to quiet somewhat when the stimulation was present, this general observation was not confirmed by an examination of the filmstrips.

Sex differences. Some variation between the behavior of males and the behavior of females was noted in each group but no sex differences were found to be both statistically significant and replicated.

Discussion

Two of the measures of newborn behavior under study, movement and hand-mouth contacting, met reasonable demands for reliability of observation and for sensitivity to individual variation. The other measures, mouthing and crying, can be observed in the neonate with satisfactory precision, but did not show in the setting described here significant sensitivity to individual differences. Of the four, movement showed systematic changes over the lying-in period.

The finding of stable individual differences in movement and hand-mouth contacting among human newborns confirms the often stated but infrequently demonstrated proposition that human beings show substantial variations one from another aboriginally. Even for the relatively homogeneous group of children observed for this research, wide and stable differences were shown. Knowing this, it is important to turn to two further questions. Do these early differences make any difference later? Does it make good sense to consider these variations, whether from time to time or from baby to baby, as representing differences in tensions? The first of these questions requires longitudinal study; the second demands an examination in some detail of the relation between activity and contacting on one hand and environmental events and other neonatal responses on the other. If neonatal movement is to serve as an index or partial interpretation of the term "tension," then it should be possible to demonstrate systematic relations to privation, to noxious external stimulation, and to the environmental operations used by parents and clinical practitioners to reduce distress in infants. On the basis of the present study, we can proceed to an attack on these next empirical questions with some confidence in the precision and the stability of measures for hand-mouth contacting and infant movement.

Summary

An initial attack on the problems of newborn behavior was made in an examination of the reliability, individual stability, and relatedness of four response measures, movement, crying, mouthing, and hand-mouth contact-

ing. Infants were seen repeatedly during the lying-in period. All measures showed very high intersubjective reliability. The measures for movement and hand-mouth contacting were sensitive to individual variation over the first five days of life; the measures for crying and mouthing were not. The movement measure showed systematic change over the lying-in period; for the other measures, neither a regular increase nor a regular decrease in occurrence was found, putting in some doubt the suggestion that mouthing and hand-mouth contacting are phenomena secondary to feeding.

The relation of the findings to the problem of tension and tension-control in the newborn was discussed.

References

1. Balint, M. Individual differences of behavior in early infancy and an objective method for recording them. *J. genet. Psychol.*, 1948, 73, 57-59; 81-117.
2. Dennis, W. A description and classification of the responses of the newborn infant. *Psychol. Bull.*, 1934, 31, 5-22.
3. Freud, S. *Three essays on the theory of sexuality.* London: Allen & Unwin, 1905 [1949].
4. Freud, S. *An outline of psychoanalysis.* Norton, 1949.
5. Fries, M. E. Psychosomatic relationships between mother and infant. *Psychosom. Med.*, 1944, 6, 159-162.
6. Fries, M. E. Diagnosing the child's adjustment through age-level tests. *Psychoanal. Rev.*, 1947, 34, 1-31.
7. Fries, M. E., & Lewi, B. Interrelated factors in development. *Amer. J. Orthopsychiat.*, 1938, 8, 726-752.
8. Gardner, G. E. Problems of early infancy. *J. Amer. Psychoanal. Ass.*, 1955, 3, 506-514.
9. Gesell, A. *The embryology of behavior.* Harper, 1945.
10. Gesell, A., & Halverson, H. M. The daily maturation of infant behavior. *J. genet. Psychol.*, 1942, 61, 3-32.
11. Graham, F. K., Matarazzo, R. G., & Caldwell, B. M. Behavioral differences between normal and traumatized newborns. *Psychol. Monogr.*, 1956, 70, No. 20 and No. 21.
12. Greenacre, P. The biologic economy of birth. *Psychoanal. Study Child*, 1945, 1, 31-52.
13. Halverson, H. M. Studies of the grasping responses of early infancy. *J. genet. Psychol.*, 1937, 51, 371-449.
14. Hoffer, W. Mouth, hand and ego-integration. *Psychoanal. Study Child*, 1949, 3/4, 49-56.

15. Irwin, O. C. The amount and nature of activities of new-born infants under constant external stimulating conditions during the first ten days of life. *Genet. Psychol. Monogr.*, 1930, 8, 1-92.

16. Irwin, O. C. The distribution of the amount of motility in young infants between two nursing periods. *J. comp. Psychol.*, 1932, 14, 429-445.

17. Jensen, K. Differential reactions to taste and temperature stimuli in newborn infants. *Genet. Psychol. Monogr.*, 1932, 12, 363-479.

18. Kendall, M. G. *Rank correlation methods.* London: Griffin, 1948.

19. Kendler, H. H. "What is learned?"—A theoretical blind alley. *Psychol. Rev.*, 1952, 59, 269-277.

20. Kessen, W., Hendry, A. L. S., & Leutzendorff, A. M. Measurement of movement in the human newborn: a new technique. *Child Develpm.*, 1961, 32, 95-105.

21. Leitch, M., & Escalona, S. K. The reaction of infants to stress. *Psychoanal. Study Child,* 1949, 3/4, 121-140.

22. Logan, F. A. *Incentive.* Yale Univer. Press, 1960.

23. Lustman, S. L. Rudiments of the ego. *Psychoanal. Study Child,* 1956, 11, 89-98.

24. Marquis, D. P. Can conditioned responses be established in the newborn infant? *J. genet. Psychol.*, 1931, 39, 479-492.

25. Miller, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* Wiley, 1951. Pp. 435-472.

26. Piaget, J. *The origins of intelligence in children.* International Universities Press, 1952.

27. Pratt, K. C., Nelson, A. K., & Sun, K. H. The behavior of the newborn infant. *Ohio State Univer. Stud., Contr. Psychol.*, 1930, No. 10.

28. Spitz, R. A., & Wolf, K. M. Autoerotism. *Psychoanal. Study Child,* 1949, 3/4, 85-120.

29. Weiss, A. P. The measurement of infant behavior. *Psychol. Rev.*, 1929, 36, 453-471.

30. Wolf, K. M. *The origins of individuality.* Mimeo., 1957.

31. Wolff, P. H. Observations on newborn infants. *Psychosom. Med.*, 1959, 21, 110-118.

FORM, COLOR, AND SIZE IN CHILDREN'S CONCEPTUAL BEHAVIOR [1]

Jerome Kagan* *and* Judith Lemkin
Fels Research Institute

Although conceptual behavior of adult humans has always been a frequent problem for investigation, studies of conceptual processes in the child have attracted considerably less attention. The work of Piaget (4) has recently catalyzed a vigor in this area and has made psychologists aware of the lack of reliable empirical information available on children's conceptual behavior. This very brief note is the first in a series of studies which seeks to understand the way in which the child conceptualizes his environment. This particular research focuses on the tendency to use form, color, or size in classifying stimuli into conceptually similar groups.

Over 30 years ago, Brian and Goodenough (1) found that the child under 3 years of age preferred form to color in organizing his world into groupings of "similar" stimuli. From 3 to 6 years of age, the child shifted to color over form as the preferred basis for categorization. After 6, form once again became dominant over color. The recent interest in conceptual behavior is flavored with a concern for individual differences rather than statements about children in general. We are now more suspicious of statements which purport to hold for all children and expect that there will be important sex, age, and even personality differences in conceptual behavior. Although Brian and Goodenough reported no significant sex differences, a recent investigation (3) with children ages 7 to 11 found that more girls than boys used form rather than color as a basis for classifying stimuli into conceptual classes. The present study is, in part, a replication of this experiment with a younger group of children.

METHOD

The subjects (*Ss*) were 35 girls and 34 boys ranging in age from 3-9 to 8-6 with a median age of 5-8. All *Ss* were enrolled in either a nursery or the first grade of an elementary school. The junior author was the experimenter (*E*), and she initially spent several hours in the classes of the *Ss*

---

* Fels Research Institute, Yellow Springs, Ohio.

[1] This research was supported, in part, by research grant (M-1260) from the National Institute of Mental Health, United States Public Health Service.

Child Develpm., 1961, 32, 25-28.

in order to promote familiarity and facilitate rapport for the individual testing situation which was to follow. Each *S* was tested individually in a room provided by the school. The *E* spent a few minutes gaining the child's confidence and then initiated the concept formation task.

Each of the nine stimuli in the series consisted of pieces of white paper upon which were pasted paper cutouts of geometrical forms differing in color, shape, and size.[2] The first three stimuli had two comparison forms on the top of the paper and a standard stimulus at the bottom. Stimuli 4 through 9 had three comparison forms on the top of the paper and the one standard stimulus at the bottom. The method used here was essentially similar to the procedure of Brian and Goodenough. For stimuli 1 through 3 the instructions were: "Here are some cutouts. Do you see this cutout down here? Which one up here (*E* points to the top of the page) is the same as the one down here?" For items 4 through 9 the instructions were: "Now we have one cutout down here and three up here. Which one up here is the same as the one down here?"

Each of the nine stimuli contained forms which could be classified with the standard stimuli on the basis of form, color, or size. For each stimulus only two of these three alternatives were possible as correct solutions. For stimulus 1, color was opposed to form; for stimulus 2, color vs. size; stimulus 3, size vs. form; stimulus 4, color vs. form; stimulus 5, form vs. size; stimulus 6, color vs. form; stimulus 7, form vs. size; stimulus 8, color vs. size; stimulus 9, color vs. size. The *E* recorded the child's answer and asked why the child had chosen that stimulus. All the children understood the task and performed appropriately, but usually were unable to answer why they had made the response they did.

RESULTS

Table 1 presents the percentage of boys and girls who chose form, color, or size for each of the nine stimuli. All children strongly preferred form to either color or size, and size was rarely used as a basis for conceptual similarity. For any one stimulus, there were no statistically significant sex or age differences in choice of color, form, or size. However, when the total number of form, color, and size responses were summed for all nine stimuli, two interesting results emerged.

Table 2 presents the average number of form, color, and size choices for boys and girls who were divided at the median into young and old groups. There was no difference between the responses of young vs. old boys. However, older girls were much less likely than younger girls to choose color ($p < .02$; two tails). Secondly, older boys were much more likely than older girls to use color as the basis for conceptualization ($p < .05$; two tails).

---

[2] The standard stimuli sampled one of three forms (triangle, square, or circle), one of four colors (orange, green, red or blue) and one of two sizes (large or small). The area of the large forms was approximately six times the area of the smaller forms.

TABLE 1

PERCENTAGE OF BOYS AND GIRLS CHOOSING FORM, COLOR, OR SIZE AS BASIS OF CLASSIFICATION

Stimulus	*Alternative*	*Preferred Choices*	*Boys*	*Girls*
1	F vs. C	F	64.7	85.7
2	C vs. S	C	73.6	60.0
3	F vs. S	F	73.6	91.4
4	F vs. C	F	64.7	80.0
5	F vs. S	F	67.7	80.0
6	F vs. C	F	58.8	77.3
7	F vs. S	F	73.6	88.6
8	C vs. S	C	53.0	51.3
9	C vs. S	C	64.7	68.5

This finding agrees with the results of Honkavaara who found that more girls than boys (aged 7 to 11) preferred form to color as a basis for conceptualization. The fact that color is not a preferred hypothesis for children this age is supported by the unexpected finding that, when color and size were present as the only bases for similarity, over 20 per cent of the *S*s said they did not know what the answer was and could not offer an answer to *E*. That is, when the stimuli did not allow for the use of form but did allow for the use of color, one fifth of the children did not know how to respond. This indicates that a color hypothesis is not merely weaker than form but, for some *S*s, is unavailable when the child is asked the question, "How are these stimuli the same?"

TABLE 2

MEAN NUMBER OF FORM, COLOR, AND SIZE CHOICES FOR ALL NINE STIMULI BY SEX AND AGE

	BOYS		GIRLS	
	Young	*Old*	*Young*	*Old*
Form	4.2	4.3	4.3	5.7
Color	3.4	3.6	3.4	1.6
Size	0.4	0.9	0.6	1.2

DISCUSSION

The results indicate that the preschool child's understanding of the phrase "same as," in the present context, is influenced primarily by the shape or form of the stimulus rather than its color or size.

The sex differences between older boys and girls is difficult to explain. One post hoc argument is presented. Since the verbal skills of young girls proceed precociously in relation to boys, it is possible that the girls implicitly apply the language labels, square, triangle, circle, to the stimuli more often than boys. Thus, for the girls, the stimuli are more likely to derive their meaning from the label attached to them rather than through the more direct physical quality of color. This explanation agrees with Sigel's (5) finding that, when school age children were asked to sort diverse stimuli into conceptually similar groups, the semantic meaning attached to the stimuli crowded out other bases for conceptualization (color, type of material, texture). Sigel concluded that the label attached to the stimulus is the primary basis for conceptualization in school age children. Brian and Goodenough reported that the brighter children were more likely to prefer form to color, and this incidental datum supports the present post hoc argument.

SUMMARY

This paper describes an experiment on the conceptual preferences of children when they were asked to group geometric stimuli differing in form, color, or size into similar pairs. Subjects, 35 girls and 34 boys (ages 3-9 to 8-6), were presented with nine stimuli in which a standard stimulus was presented simultaneously with other stimuli differing in form, color, and size. The child was asked to select the comparison stimulus which was the "same as" the standard. The result indicated that for both boys and girls form was distinctively preferred to color as a basis for similarity and that color was preferred over size. For boys, there was no age difference in this response pattern. However, older girls were less likely than younger girls to use color as a basis for conceptualization. Older boys, moreover, were more likely to use color than were the older girls. These results agreed with a recent investigation in this area, and it was suggested that the sex difference was a function of the girls' implicit labeling of the forms as square, triangle, or circle, which presumably facilitated the use of form as a basis for similarity.

REFERENCES

1. BRIAN, C. R., & GOODENOUGH, F. L. The relative potency of color and form perception at various ages. *J. exp. Psychol.,* 1929, 12, 197-213.
2. BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking.* Wiley, 1956.
3. HONKAVAARA, S. A critical reevaluation of the color and form reaction, and disproving of the hypotheses connected with it. *J. Psychol.,* 1958, 45, 25-36.
4. PIAGET, J. *The construction of reality in the child.* Basic Books, 1954.
5. SIGEL, I. E. The dominance of meaning. *J. genet. Psychol.,* 1954, 5, 201-207.

DEVELOPMENTAL CHANGES IN PREFERENCE FOR GOALS DIFFICULT TO ATTAIN[1]

Donald C. Smith* *and* Lucy Wing**

Syracuse University

In the 1930's Wright (3) conducted a series of experiments with children and adults to test the hypothesis that in a choice situation an obstructed goal has a stronger positive valence than an accessible one. His findings, which generally supported the hypothesis, also indicated that preference for a difficult goal is positively associated with increasing age. Adults in his study more consistently chose the difficult goal than did children. Preferences were also found to vary with the degree of difficulty of the barrier, with the nature of the goal object (*a goal in itself* or *a means to an end*), and with manner of expression of choice (verbally or through action).

Child and Adelsheim (2) replicated Wright's experiments with slight variations. They found that nursery school children consistently preferred the nearer, more available goal. A later developmental study by Child (1) indicated that preference for the nearer of two goals was much more frequent among younger children than among older children.

The present experiment is an extension of Child's developmental study. One of the acknowledged limitations of his research was the use of only one type of barrier. The present research employed a barrier of difficulty

---

* Bureau of Educational Research and Service, Ohio State University, Columbus 10.

** Department of Psychology, Oswego State Teachers College, Oswego, New York.

[1] The writers are greatly indebted to Dr. George G. Thompson for valuable advice and criticism. They also wish to thank Principal Hilda Jacobson and Principal John Dunn of the Liverpool School System for their cooperation in making available subjects for the study.

Child Develpm., 1961, 32, 29-36.

as well as a barrier of distance. Goal objects were employed which were classified as goals in themselves and as means to ends. An attempt was also made to control more rigorously the following variables which previous experiments had found related to preferences for more difficult goals: sex, intelligence, socioeconomic status, racial origin, and academic achievement of subjects, and sex of the experimenters. As additional controls, an equal number of goal and means objects were included, barriers were selected which were well within the capacity of the youngest child, and communication was prevented between subjects prior to testing.

The main purpose of the present experiment was to investigate developmental changes in preference for goals which are more difficult to attain. The experimental hypothesis was that preference for difficult goals increases with increasing chronological age. It was assumed that, through experiences of social approval, children in our society learn to place a high value on the achievement of difficult things. As they grow older, and there are more occasions for social reinforcement of their preferences, the frequency of choice of difficult goals was expected to increase.

Three other questions were considered in the analysis of preferences for difficult goals: (a) the effect of the sex of the examiner, (b) the effect of the nature of the task, (c) the effect of goal objects which are ends in themselves or means to an end.

SUBJECTS

The subjects were 80 boys enrolled in the Liverpool, New York Public Schools. Groups of 20 boys from grades 2, 4, 6, and 8 were closely equated on intelligence and socioeconomic status. Intelligence test scores from the California Test of Mental Maturity were used. The mean intelligence quotients (IQs) for grades 2, 4, 6, and 8 were 118, 114, 114, and 115, respectively. IQs for the total group ranged from 94 to 144. Socioeconomic status was based on the occupations of fathers of *S*s, rated according to the University of Chicago revision of the Warner, Meeker, and Eells Scale for Rating Occupations (4). Occupational ratings of fathers ranged from 2 through 5 on the Index of Status Characteristics, and the mean occupational ratings were identical for the four grade levels. The average socioeconomic ratings fell in the lower middle-class range.

Only white, native-born *S*s were included and only those whose school achievement was within the average to above-average range on the basis of standardized achievement tests and judgments of teachers. Any child who in the opinion of the teacher, guidance counselor, or the helping teacher was physically handicapped or seriously emotionally disturbed was excluded from the study. The age range for the total group was 7-7 to 14-4 years. Mean ages for grades 2 through 8 were 8-1, 10-0, 11-8, and 13-9, respectively. Standard deviations in months for the four grade levels were 3.0, 4.1, 6.6, and 3.2.

EXPERIMENTAL PROCEDURES

Subjects were randomly assigned to the male and female experimenters so that 10 boys at each grade level were seen by each. Boys were taken from their classrooms one at a time and followed each other in rapid succession. While one subject was being seen, another was taken to the experimental room. Schedules were carefully checked so that playground, lunch, or other free periods did not intervene before all subjects in one classroom had completed the experiment; therefore, subjects already taken from the home room had no opportunity to discuss the situation with future subjects.

After personal data had been secured, the experimenter explained to each subject that the purpose of the study was to find out the kinds of things boys his age liked to do and emphasized that it had nothing to do with school work. Previously, teachers had been instructed to underplay the situation, and no undue curiosity was aroused. Most of the children were accustomed to occasional withdrawal from their classrooms for individual testing.

The four experimental tasks placed each subject in the position of having to choose between two goals, one accessible and the other relatively obstructed. The obstructed goals involved the barriers of greater distance or greater difficulty. The Table and Ladder situations, which represented the barrier of distance, were identical to the experimental tasks used by Child (1). On the Formboard and Assembly, the barrier was considered to be the number of puzzle pieces available for completion of the tasks. For the former two situations, candy served as the goal object; for the latter two situations, the optional number of puzzle pieces were considered to be *means to an end* goal objects.

Table Situation. *S* was asked to stand at one end of a table. A piece of chocolate-covered cream candy was placed directly in front of him and an identical piece of candy was located at the center of the opposite end of the table. The boy could reach the nearer candy easily but had to walk around the table to secure the more distant piece. Both pieces of candy were wrapped in cellophane so there would be no rejection for sanitary reasons. *E* said, "Now would you please stand here at the end of the table? Do you see the two pieces of candy? You may have either one you want." If *S* asked if the two pieces were alike, a noncommittal reply was given, "Well, I don't know. They look about the same."

Ladder Situation. Next, *S* was asked to stand by a three-step ladder which was in front of a book shelf. One piece of chocolate candy was placed within easy reach on a low shelf. The other candy was put on a high shelf so that *S* had to climb the steps to secure the more distant goal. The placement of the candy on the shelves was adjusted to the height of subjects so that the distance would be the same for both younger and older children. *E* said, "Now, come and stand over here" (indicating a place directly in front of the shelves). "Do you see those two pieces of candy? You could get one by climbing the steps. You could get this one by reaching right over here. You may take either one you want."

Formboard Situation. The equipment was a nine-inch formboard, and subjects had the choice of completing this with either eight or 16 triangular pieces. The pieces were stacked in two separate piles; whether the pile with the most pieces was to the right or left of *S* was predetermined by random assignment. *E* said, "Now, let's sit down here at the table. Here is a puzzle we would like you to try. The pieces in both piles will fit into this space. This pile has more pieces than the other one. I'd like you to finish the puzzle by using the pieces in one of these piles. You may choose either one of the piles you wish. Which one do you want to use? (Pause) All right, put the puzzle together." *E* did not show interest in a subject's proficiency in doing the puzzle. If a child had much difficulty and showed any signs of frustration, *E* helped him to complete the formboard.

Assembly Situation. This task involved a simple serial-order assembly consisting of wooden forms which fitted into a grooved base. The subject could select an assembly with 10 or 15 pieces. The size and shape of these pieces were identical in both cases. The following instructions were given: "Here's another kind of puzzle. What you do is take all the pieces out and then put them back together. This one has more pieces in it than the other one. You may choose either one. Which one do you want to do? (Pause) All right. Go ahead and do it." The 10- and 15-piece assemblies were randomly assigned to the right or left side of subjects.

RESULTS

The main purpose of the experiment was to establish whether choices of more difficult goals increased proportionally with age. This first hypothesis was not supported, as shown in Figure 1. The mean percentage of children choosing the more difficult goals increased from 32 per cent at

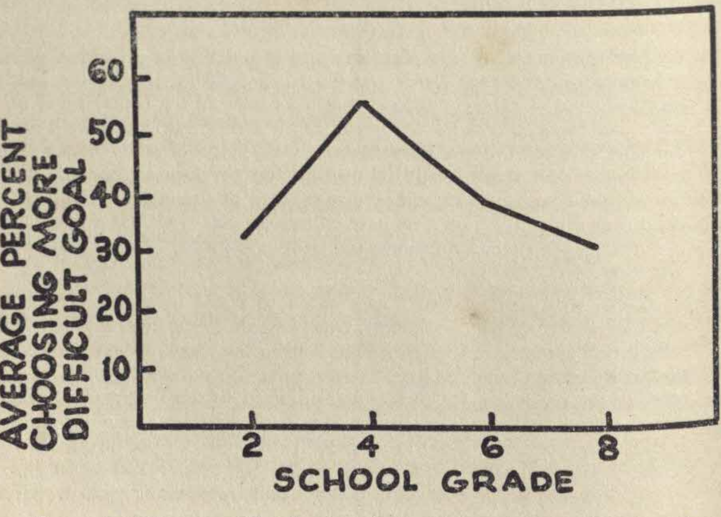

FIGURE 1—Changes with school grade in choice of more difficult goal.

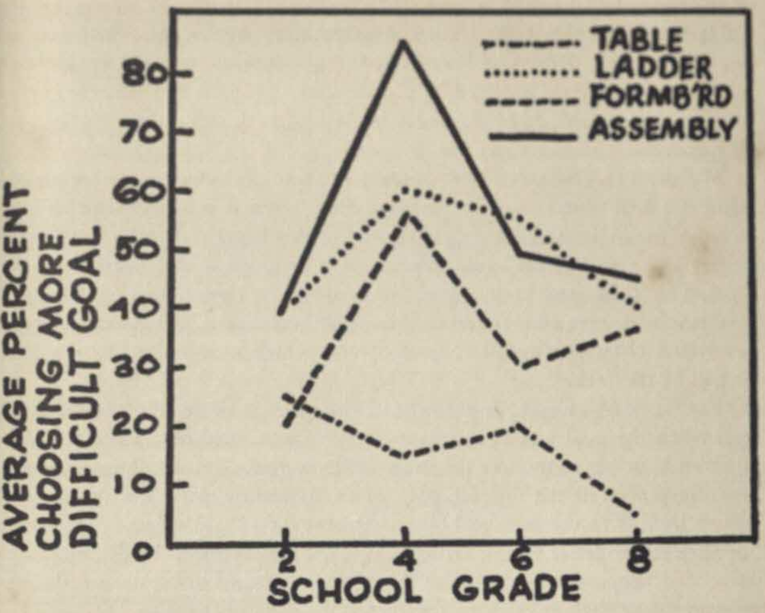

FIGURE 2—Changes with school grade in choice of more difficult goal in the four different situations.

grade 2 to 55 per cent at grade 4 but decreased to 39 per cent and 31 per cent at grades 6 and 8. Thus, when responses to all four tasks are combined, the percentage of boys choosing the more difficult goal was in the direction predicted for grades 2 and 4 but decreased progressively for grades 6 and 8.

As may be seen in Figure 2, slight variations occurred in the cases of the Table situation and the Formboard task. The marked increase at the fourth-grade level was found with all the experimental tasks except for the Table situation. Although the general trend is downward from the fourth grade on, slight deviations may also be noted on the Formboard and Table situations.

Child (1) reported that the sex of the experimenter was a variable which had an important bearing on the responses of subjects. His experiment indicated that subjects were more likely to choose the more difficult goal with a female than with a male experimenter. However, in the present study, no significant differences were found between the responses of the subjects who worked with the male experimenter and those who worked with the female experimenter.

The second experimental variable investigated was the influence of the nature of the task upon preference for more difficult goals. It was suspected that preference for the more difficult goal would vary considerably in terms of the type of task which confronted the child. Chi square values for the Table and Assembly tasks were not statistically significant. However, on two of the tasks, Table and Formboard, fewer subjects selected the difficult goal than would have occurred by chance alone ($p < .01$). Thus, the nature of the task appeared to be definitely associated with the preference for specific goals.

Wright (3) concluded that difficult goals are selected more frequently when the goal object is an end in itself than when it is a means to an end. In order to obtain additional data related to this hypothesis, two "end" goal objects and two "means" goal objects were included in the present experiment. The Table and Ladder situations involved a choice between two goal objects which were *ends* in themselves. The Formboard and Assembly tasks involved a choice between two goal objects which were *means* to the completion of the task.

For the total sample, 32 per cent of the subjects chose the more difficult goal when the goal was an end in itself. When the goal object was a means to an end, 45 per cent chose the more difficult goal. Almost all subjects consistently preferred the difficult goal more frequently with the *means* goal objects than with the *ends* goal objects, except for a slight negative difference for subjects in grade 2. The critical ratio test was applied for differences in correlated proportions, and the differences were found to be statistically significant ($p < .05$). The data, therefore, are contradictory to Wright's hypothesis. It could be legitimately argued, however, that the greater preference for the difficult goals on the Formboard and Assembly task was due to the difference in types of barriers rather than to the difference in goal objects.

Discussion

Despite differences in sampling and in experimental tasks, the present findings are in close agreement with those obtained by Child (1). He found that about 20 per cent of first and second graders chose the more distant or more difficult goal. In grades 3 and 4 this proportion increased to slightly above 50 per cent. After this maximum had been reached, the proportion tended to fluctuate irregularly in succeeding years. Thus, in Child's study, the preference for the more difficult goals increased to a point somewhere above eight or nine years and fluctuated irregularly thereafter. In the present experiment, preference for the more "distant" or more "difficult" goals increased to a maximum of 55 per cent at grade 4 (approximately age 10) and decreased progressively thereafter.

Child questioned his results at the upper grade levels. He reasoned that flaws in procedures caused the irregular fluctuation. Older children had the opportunity to communicate freely among themselves prior to entering the experimental situation. They talked about the purpose of the experiment and about the appropriate response. Because of this communication among *S*s, their choices were considered to be group rather than individual decisions.

Thus, the irregularity of choices in Child's study at the upper grade levels could have resulted from this situation rather than from a basic irregularity in preference for more difficult goals.

In the present experiment, careful control was exercised over communication among subjects. No subject in any grade had the opportunity prior to being tested of discussing the experiment with another member of his grade. Thus, choices were independently and individually determined. With this factor controlled, irregularity in preference for difficult goals was not found at the upper grade levels, neither when Child's two tasks nor when all four tasks were considered.

It seems probable that preference for a difficult goal is not a simple generalized trait at any age level. Choice of any individual may be influenced by a multiplicity of variables: How highly is the goal object valued? How strong is the barrier to the achievement of the goal and what form does the barrier take? What situational forces are acting upon the child? What is the sequence of activity? How does the child express his preference? Different patterns of interaction between these variables may exist at different age levels.

SUMMARY

The main question under investigation was whether preference for the goal more difficult to attain is a generalized trait which increases in proportion to increasing age. Secondary questions were whether the sex of the experimenter and the type of barrier had a significant effect on preference.

Groups of 20 boys from grades 2, 4, 6, and 8 were closely equated on intelligence and socioeconomic status. Boys were assigned randomly to experimenters at each grade level. Each subject was confronted with four tasks in which he was asked to make a choice between a goal difficult or easy to attain. The barriers to the attainment of goals consisted of distance between the subject and the goal and of the number of pieces used to complete Formboard and Assembly tasks.

The first hypothesis was not supported. The percentage of boys choosing difficult goals under these experimental conditions increased from 32 per cent at grade 2 to 55 per cent at grade 4 but decreased to 39 per cent and 31 per cent at grades 6 and 8. No significant differences were found in preference for difficult goals between the individuals assigned to the male or female experimenter. The nature of the task and the type and strength of barrier used had a significant effect on choices of difficult goals.

This experiment suggests that preference for difficult goals is not a simple generalized trait which increases with age. At any age level preferences may vary in terms of a number of associated variables. Important factors are the intrinsic value of the goal object, the strength of the barrier, the motivating conditions within the experimental situation, the sequence of activity, and the manner in which choice is expressed by the child. Subsequent research

should consider the interaction of these variables and their influence on preference for easy- or difficult-to-attain goals.

References

1. Child, I. L. Children's preferences for goals easy or difficult to obtain. *Psychol. Monogr.*, 1946, 60, No. 4 (Whole No. 280).
2. Child, I. L., & Adelsheim, E. The motivational value of barriers for young children. *J. gen. Psychol.*, 1944, 65, 97-111.
3. Wright, H. F. The effect of barriers upon strength of motivation. In R. G. Barker, J. S. Kounin, & H. F. Wright (Eds.), *Child behavior and development.* McGraw-Hill, 1943. Pp. 379-396. (Originally published in a monograph in 1937.)
4. University of Chicago. Methodological and operational considerations in the application of the Index of Status Characteristics to the metropolitan community, 1956.

EFFECT OF EXPOSURE TO SYMBOLIC AGGRESSION ON AGGRESSIVE BEHAVIOR[1]

O. Ivar Lövaas* [2]
University of Washington

This study was concerned with the effect of exposure to symbolic aggression on the play behavior of children. A bar-pressing response which produced aggressive doll action was observed immediately after the children had been exposed to either an aggressive or a nonaggressive motion picture.

Previous research on changes in aggressive behavior has been oriented primarily around testing the hypothesis of a cathartic or drive-reducing effect of aggressive expression (2). The data tend to support the hypothesis of drive-reduction when the aggressive expressions have been varied in the immediate experimental situation. Feshbach (5) observed a decrease in aggressive behavior immediately following the writing of aggressive TAT stories. Similarly Lewin *et al.* (10) observed a reduction in the over-all level of aggression in children subsequent to scapegoating behavior. When the experimental condition consisted of more passive exposure to aggressive behavior, such as viewing aggressive films, the reduction in aggressive behavior has not been demonstrated. Siegel (11) failed to find a difference in amount of aggressive play behavior in children who had been exposed to an aggressive film. This difference in findings is analogous to results concerning behavior for reinforcers based on the primary appetitive drives. If the organism does not consume foodstuff, but merely tastes it or sees it, his behavior to obtain foodstuff is increased rather than decreased (1, 7).

The assumption is made that the aggressive stimuli (of doll action) act as reinforcing stimuli. It is supported by Burns' data[3] which show that, when an aggressive action (in dolls) was contingent upon a response (lever-press), that response was strengthened. The remaining conceptual frame-

---

* Gatzert Institute of Child Development, Department of Psychology, University of Washington, Seattle 5.

[1] This study was supported by a grant from the National Institute of Health (M-2208).

[2] The author expresses his gratitude to Professors Donald M. Baer and Sidney W. Bijou for their continued assistance in this research. He also is indebted to staffs at the Seattle Day Care Nursery and at the Nursery School in the Gatzert Institute of Child Development, University of Washington, for their cooperation.

[3] Personal communication from R. C. Burns.

Child Develpm., 1961, 32, 37-44.

work for the study was oriented around research in relation to primary appetitive reinforcers. An attempt was made to investigate whether there is a change in aggressive behavior following exposure to aggressive stimuli analogous to the change in eating behavior following exposure to foodstuff. A similar conceptual framework was proposed by Gewirtz and Baer (6) in their study on the reinforcing properties of social approval. Their study demonstrated that the effectiveness of social approval as a reinforcer was related to satiation and deprivation operations in the manner analogous to such operations on primary appetitive reinforcers.

EXPERIMENT I

Method

The research was conducted in the Gatzert Institute of Child Development. Two experimental rooms were employed. The "playroom" contained the play apparatus, a screen for the presentation of the films, and a partition in one corner behind which *A* (an adult assistant) could sit. The other room, an observation room with one-way mirrors and sound equipment, contained the apparatus for presentation of the films and recording of behavior. The *E* (experimenter) could present the various experimental treatments from this room without being seen by the *S* (subject).

The films were presented by a Busch "Cinesalesman Continuous Projector" and appeared on a 9 by 12 inch screen in one wall of the playroom. A shutter blocked the lens every 10 seconds, for a period of five seconds. The *S*s were told that they could turn the film on when it went off by depressing a four-inch lever protruding from a small box on a table in front of the child. If the lever was pressed any time during the five-second off period, the shutter would reopen and allow for another 10 seconds of viewing before the next interruption. Having the *S*s "operate" the films in this manner assured continued attention to the films.

Two films were employed. The aggressive film (Official Films: "Rassling Match") was cut to give an almost continual display of aggression (hitting, biting, etc.) inflicted by one humanlike cartoon figure upon another. The nonaggressive film (Castle Films: "Bear Facts") depicted three bear cubs and a mommy bear engaging in pleasant humanlike play. The cut versions of each of these films had a running time of five minutes.

The doll apparatus was similar to that employed by Burns.[3] The dolls were six inches high, similar looking, dressed in boys' clothes, and facing each other about three inches apart. The doll on the right held a stick in its hand. They were placed inside a plexiglass box on top of a wooden box with a four-inch lever protruding from the front. Depression of this lever would activate a concealed motor which would swing the doll's arm with the stick rapidly down and back up into its original upright posititíon, making the stick hit the other doll on the head. Operation of the lever also activated an electric counter.

The subjects were children from the Institute's Nursery School. They were above average in intelligence and came from families with fathers in occupations above the means of the general population (3). Twelve children from the five-year-old group were selected in a random manner except for elimination of those whose cooperation was known to be difficult to obtain. They had considerable experience as *S*s in other experiments, but had no previous contact with the films or dolls employed in this study. Except for matching on sex, they were randomly divided into two groups of six each. One group was exposed to the aggressive film (AF group), the other group to the nonaggressive film (NAF group); otherwise, they were given the same experimental treatment. The *S*s were seen in the morning while they were engaged in free play. Both groups were run on any one day and matched on hour of the day. They were seen for three trials, spaced about four days apart.

A, a female adult selected because she was "good" with children, had visited the nursery prior to the experimental run and acquainted herself with the children. Each *S* was accompanied to the playroom by *A* who would invite the *S* to "play some games and look at a movie." (The playroom was located in the same building as the nursery.) When the *S* entered the playroom, he was made familiar with the room first and then introduced to the dolls. *A* pressed the lever once and said, "When you press this bar this (pointing) will happen. Now you sit in your chair and play with this while I sit in my chair." He received a two-minute period of continual reinforcement, i.e., one doll action for each bar press. The period was timed from the *S*'s initial depression of the lever. This period constituted the pre-experimental or operant level of doll action. *A* then placed the dolls out of the *S*'s view and said, "You will now see a movie. It will be shown right here (pointing to the screen). You are going to be the operator. That means that you have to turn the movie on when it goes off. Each time the movie goes off, press this lever (pointing to the box on the table) and the movie will come back on." The film was started and *A* made sure that the *S* knew how to turn the film on. The film continued for five minutes after which *A* told the *S* to play some more with the dolls. A two-minute period of continuous reinforcement followed, timed from the initial bar-press after the film. At the end of the session *S* received a "thank you" and was accompanied back to the nursery.

A interacted with *S* in a matter-of-fact manner. During the session she remained behind a partition except when she introduced the film or the dolls. She did not respond to *S*'s attempt to converse, unless he came over to her chair at which time she stated: "I'll sit here and you be over there for a few minutes more."

The second and third trials were identical with the first, with two exceptions. First, the operant period was eliminated. Second, doll play in the third period was immediately followed by a one-minute extinction period in which the second, fourth, and eighth response were reinforced to assure prolonged extinction. One *S* from each of the groups refused to return for the third trial. In an attempt to prevent further loss of *Ss*, the reinforced doll play session in the third trial was shortened to one minute.

Results and Discussion

The differences in the mean number of responses on the dolls after exposure to the films were not statistically significant. The AF group showed fewer responses than the NAF group in the first postfilm session. This difference reversed itself by the last session (the extinction session). A second experiment with a larger *N* was undertaken in an attempt to obtain a statistically significant difference between the groups in the first postfilm session.

EXPERIMENT II

Method

This experiment was conducted in a trailer laboratory (4). The trailer was parked in the play yard of a day care center supported by the Community Chest. This school accepted children only from low income families with working mothers. The *Ss* were 20 children from four to six years of age. Except for matching on sex and age, they were randomly divided into two groups of 10 *Ss* each, forming AF and NAF groups analogous to those in Experiment I. None of the children had previous familiarity with the films or the dolls. The experimental procedure was identical to that in Experiment I, except that a four-minute extinction period followed immediately after the two-minute postfilm session of continual reinforcement. The experimental treatment was: a two-minute continual reinforcement, operant session; five minutes of the respective film; a two-minute period of continual reinforcement; a four-minute period of extinction with the second, fourth, and eighth response reinforced.

Results and Discussion

Despite the increase in *N*, none of the differences between the groups was statistically significant.

At the completion of Experiment II the experimental conditions were re-examined for possible sources of interference with the behavior following seeing the film. There was a possibility that the behavior involved in complying with instructions was stronger than the behavior resulting from the experimental conditions. In the preceding studies it appeared that some of the *Ss* persisted in pressing the lever for the doll action because they were told to do so by *A*. Many *Ss* showed signs of uneasiness about leaving the dolls during the latter part of the period, but they would continue pressing the lever without even attending to the dolls. The verbal instructions apparently kept the *Ss* operating the dolls, hence competing with and obscuring the effect of the films in controlling this behavior. This notion was investigated in Experiment III.

EXPERIMENT III

Method

The experimental setup was modified to minimize the interfering effect of "following instruction." Another toy was placed in the situation giving the *S* a choice for responding. This toy consisted of a lever and a wooden ball enclosed in a cage. Depression of the lever made the ball go to the top of the cage and back, through obstacles, to its original position. Depression of the lever also activated an electric counter. The boxes containing the dolls and the ball toy were placed side by side on a table, enabling *S* to work both simultaneously if he so desired.

The same *S*s used in Experiment II returned a week later but with reversed conditions, i.e., the group who had seen the aggressive film in Experiment II saw the nonaggressive film in Experiment III and vice versa. Any difference between the two groups in the response rate on the dolls would have to be attributed to the films, as operant level of responding was identical for the two groups.

When *S* came back to the trailer this time, he was introduced to the ball toy, shown how it worked, and asked to play with it for one minute. He was then reintroduced to the dolls and asked to press the bar to see if he remembered how it worked. He was allowed to press it five times and was then presented with the (different) film. After film viewing was completed, he was told, "You may play with this (pointing to the ball toy) or this (pointing to the dolls and lever), play with the one you like but start playing with this (pointing to the dolls and lever)." *S* was requested to play with the dolls first to insure that some behavior was recorded on the dolls. *A* waited beside *S* until he made his first response on the dolls, then retired behind the partition. It was intended that he should be free to leave the dolls whenever he wanted. If after two minutes he had not played with the ball toy, he was told, "You may play with either one, play with the one you want to play with." This session lasted for four minutes.

Results and Discussion

The total number of responses on the bar operating the dolls and the ball toy for the two groups are presented in Table 1 and an analysis of variance of these data is presented in Table 2.

The statistical tests show a significant interaction between the kind of toy played with (ball and doll) and the kind of film preceding the toys. When the *S*s could choose between the toys, the *S*s who had been exposed to the aggressive film engaged in more play behavior with the hitting dolls than did *S*s who had been exposed to the nonaggressive film. The AF group had almost twice as many responses on the toy with the aggressive doll action as the NAF group. This difference is statistically significant at the $p < .06$ level, one-tailed test ($t = 1.62$, $df = 18$). The variability between *S*s was considerable. Also the difference between the two groups was attenu-

TABLE 1

RESPONSES ON DOLLS AND BALL FOLLOWING AGGRESSIVE AND NONAGGRESSIVE FILMS

	MEAN RESPONSES ON TOY		
	Dolls	*Ball*	*Sum*
Aggressive Film	98.2	99.5	197.7
Nonaggressive Film	58.6	142.1	200.7
Sum	156.8	241.6	

TABLE 2

ANALYSIS OF VARIANCE FOR RESPONSES ON DOLLS AND BALL FOLLOWING AGGRESSIVE AND NONAGGRESSIVE FILMS

Source of Variation	*Sum of Squares*	*df*	*Mean Sqare*	*F*
Between Films	22.5	1	22.5	
Between Toys	21529.6	1	21529.6	5.77*
Interaction: Films × Toys	16892.1	1	16892.1	4.53*
Within	134361.8	36	3732.2	
Total	172806.0	39		

* Significant at $<$.05 point.

ated by "obedient" *S*s, i.e., *S*s who were told at the end of two minutes of continuous doll play to play with either toy, immediately switched to the ball toy and stayed with that toy for the remaining two minutes without switching back to the dolls. There were two such *S*s in the AF group, four in the NAF group. The four "obedient" *S*s in the NAF group had a higher response rate than the remaining *S*s in the NAF group. They obscured the difference between the groups by accentuating the over-all response rate of the NAF group.

The results of this study give evidence for an increase in responding to aggressive reinforcing stimuli as a function of exposure to symbolic aggression. From a theoretical viewpoint, exposure to symbolic aggression may have at least four effects upon subsequent aggressive behavior. First, the exposure may satiate the organism on aggressive reinforcers. Secondly, it may constitute an operation which increases the organism's generalized drive level (13). Thirdly, it may sensitize the organism, in the manner of Spence's *K* (13), by making behavior with aggressive consequences more probable. Fourth and lastly, the operation may provide discriminative stimuli, in the manner of Skinner's analysis (12), marking the occasion whether aggressive behavior can occur or not.

The first of these alternatives was not supported since there was an increase in behavior with aggressive consequences rather than a decrease which should occur under satiation. The second consequence, pertaining to increased generalized drive, similarly was not supported. Inspection of Table 1 indicates that the total means of the two groups on the two toys was almost identical (198 for the AF group, 201 for the NAF group). The higher response rate of the AF group on the dolls did not come about because of a generalized heightened activity level as a function of increased generalized drive, but rather because the aggressive film selectively increased behavior that had aggressive consequences.

The specific increase in responding for aggressive consequences is consistent with both alternative three and four. The third alternative, concerning increased sensitization to behavior with aggressive consequences, can be presented in Spence's framework (13) in the following manner. Availability of or exposure to a reinforcing stimulus can produce either an increase or a decrease in behavior for that reinforcer. If the exposure includes only a visual response to the reinforcing stimulus, *K* (incentive motivation) will be increased. An increase in *K* brings about an increase in probability of responding for that reinforcer.

The fourth alternative, change in the discriminative stimuli as a function of exposure to aggression, might be described in Skinner's framework (12) as follows. The sight of aggression, like the sight of food, constitutes a discriminative stimulus. For these children this particular film constituted a discriminative stimulus marking the occasion when aggressive behavior would be reinforced positively or at least not punished. A different film could mark the occasion when a child could not aggress successfully; hence it could constitute a discriminative stimulus for not aggressing.

The films employed in this study were not unlike those presented for children's entertainment, e.g., on TV, and the children did not represent an unusual sample of preschool children. There is a growing body of empirical evidence to support generalizations from play behavior to social behavior (8, 9). Hence the educational implication from this study is that aggressive films are likely to make children more aggressive rather than less aggressive.

SUMMARY

Three experiments were performed on the effect of exposure to symbolic aggression by means of motion pictures on the play behavior of children. A bar-pressing response which produced aggressive doll action was observed immediately after the children had been exposed to either an aggressive or a nonaggressive film. The study gave evidence for an increase in responding for the aggressive doll action after exposure to the aggressive film when such responding is properly observed. The data are analogous to findings concerning behavior for reinforcers based on the primary appetitive drives. The

effects of this visual exposure to symbolic aggression can be conceptualized either in terms of an increase in incentive motivation (K) or in terms of providing discriminative stimuli marking the occasion when aggressive behavior will be reinforced or at least not punished.

REFERENCES

1. ANDERSON, E. E. The externalization of drive. IV. The effect of prefeeding on maze performance of hungry non-rewarded rats. *J. comp. Psychol.,* 1941, 31, 349-352.
2. BERKOWITZ, L. The expression and reduction of hostility. *Psychol. Bull.,* 1958, 55, 257-284.
3. BIJOU, S. W. Patterns of reinforcement and resistance to extinction in young children. *Child Develpm.,* 1957, 28, 47-54.
4. BIJOU, S. W. A child study laboratory on wheels. *Child Develpm.,* 1958, 29, 425-427.
5. FESHBACH, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.,* 1955, 50, 3-11.
6. GEWIRTZ, J. L., & BAER, D. M. Deprivation and satiation of social reinforcers as drive conditions. *J. abnorm. soc. Psychol.,* 1958, 57, 165-172.
7. GRINDLEY, G. C. Experiments on the influence of the amount of reward on learning in young chickens. *Brit. J. Psychol.,* 1929, 20, 173-180.
8. HOLLENBURG, E., & SPERRY, M. Some antecedents of aggression and effects of frustration in doll play. *Personality,* 1951, 1, 32-43.
9. LEBO, D., & LEBO, E. Aggression and age in relation to verbal expression in nondirective play therapy. *Psychol. Monogr.,* 1957, 71 (20).
10. LEWIN, K., LIPPITT, R., & WHITE, R. K. Patterns of aggressive behavior in experimentally-created "social climates." *J. soc. Psychol.,* 1939, 10, 271-299.
11. SIEGEL, A. E. Film-mediated fantasy aggression and strength of aggressive drive. *Child Develpm.,* 1956, 27, 351-365.
12. SKINNER, B. F. *The behavior of organisms.* Appleton-Century-Crofts, 1938.
13. SPENCE, K. W. *Behavior theory and conditioning.* Yale Univer. Press, 1956.

SEX DIFFERENCES IN AGGRESSION AND ITS CORRELATES IN MIDDLE-CLASS ADOLESCENTS[1]

LEONARD M. LANSKY,* VAUGHN J. CRANDALL, JEROME KAGAN, *and* CHARLES T. BAKER[2]

Fels Research Institute for the Study of Human Development

Students of human development agree that adolescence is one of the crucial developmental periods. It is a time when the problems of dependence and autonomy with respect to parents and other adults are salient, and a period when sexual attitudes and social role identities are clarified. However, there is a very limited body of scientific information about adolescents, and even less rigorous theory. This paper presents some results concerning relations among a number of behavioral areas salient for normal, middle-class adolescents. The areas under study included aggression, achievement, dependence, independence, affiliation, identification, and moral standards. This report focuses on sex differences in the relation between aggressive behaviors and other behavioral areas.

METHOD

Subjects

The *S*s were 54 children between 13 and 18 years of age currently enrolled in the longitudinal study of the Fels Research Institute for the Study of Human Development. The sample was comprised of 32 boys and 22 girls. At the time of the first assessment contact, the boys ranged in age from 13-7 to 17-10; the girls, from 13-4 to 17-11. At the time of the second assessment session, the age-range of the boys was 13-10 to 18-7; of the girls, from

---

* Graduate School of Education, Harvard University, Cambridge 38, Massachusetts.

[1] This research was supported in part by research grant M-1260 from the National Institute of Mental Health, U. S. Public Health Service.

[2] Now at the University of Rochester School of Medicine.

Child Develpm., 1961, 32, 45-58.

13-5 to 18-8. The *S*s were white and came from predominantly middle-class homes. Their parents were better educated than national averages. The mean Stanford-Binet IQ of the children was 113 with a standard deviation of 11.4.[3]

Assessment Procedures

Two contacts were made with each child. The first session included a semistructured individual interview lasting between one and one-half and two hours followed by a modified Rorschach test. For the second contact, the *S*s were seen in small groups, and questionnaires and tests were administered.

Interview.[4] The interview schedule was designed to explore the following areas of behavior: achievement, dependence, affiliation, self-esteem, aggression towards parents and authorities, attitudinal differences with parents, anxiety over heterosexual relationships, and identification with parents and parental models. Three of the writers (LL, VC, JK) each interviewed approximately one third of the *S*s. Each interview was electrically recorded with the *S*'s knowledge.

Modified Rorschach. Following the interview, each *S* was administered a 32-item modified Rorschach test. The *S*'s first response and response latency were recorded for each of the 32 stimuli. The modified Rorschach stimuli were constructed by placing a template of white cardboard over part of the standard Rorschach plate leaving a common *D* or *d* area exposed. A description of the stimulus areas appears in a separate report (8).

Group administered tests. The remaining tests were administered to the *S*s in small groups (two to six subjects in a group) during the second assessment contact. Instructions to the *S*s stressed: (a) the tasks were not tests in the sense that there were any right or wrong answers; and (b) no premium would be put on speed. Since all *S*s began each test together, reading material was available for those who finished early.

In the order of administration, the tests were the Gough Brief Femininity Scale (7), the Franck Drawing Completion Test (5), the French Insight Test (6), a story completion test, and a self-rating inventory.

The Variables

The variables and instruments from which they were obtained are listed in Table 1. The variables are grouped by the behavioral areas of aggression, achievement, affiliation, dependency, sex anxiety, sex role identification, severity of standards, and defenses against guilt.

The interview ratings. Immediately following each interview, the interviewer rated the *S* with respect to 17 variables. Each scale used for these

---

[3] The IQ tests were administered by Dr. Virginia L. Nelson.

[4] A copy of the interview schedule and other assessment techniques mentioned in this report may be obtained by writing to the Department of Psychology, Fels Research Institute, Yellow Springs, Ohio.

TABLE 1

FINAL VARIABLES AND THEIR SOURCES

Variable	*Source*
Aggression	
Expressed criticism of father—I	Interview
Expressed criticism of mother—I	Interview
Reported aggression towards father—I	Interview
Reported aggression towards mother—I	Interview
Aggression—SRI	Self-rating Inventory
Aggression—R	Rorschach
Achievement	
Anxiety about lack of achievement—I	Interview
n Achievement—Fr	French Test
n Achievement-recognition—Fr	French Test
Dependence	
Conformity to authority—I	Interview
n Dependence—Fr	French Test
n Autonomy—Fr	French Test
Affiliation	
n Affiliation—Fr	French Test
Acceptance—SRI	Self-rating Inventory
Sex Anxiety	
Latency to male cards—R	Rorschach
Latency to female cards—R	Rorschach
Sex Role Identification	
Identification with father—SRI	Self-rating Inventory
Identification with mother—SRI	Self-rating Inventory
Franck (masculinity-femininity)	Franck Test
Gough (masculinity-femininity)	Gough Femininity Scale
Severity of Standards	
Severity of standards—SCB	Story Completion—Boxes
Severity of standards—SCP	Story Completion—Photography
Defenses against Guilt	
Denial SCB (SCP)*	Story Completion Item
Intellectualization SCB (SCP)*	Story Completion Item
Accept responsibility SCB (SCP)*	Story Completion Item

* Each variable was considered independently for the boxes story (SCB) and the photography story (SCP).

ratings contained an introductory description of the variable and a six-point scale with cue-point referents for the extreme points (1 and 6) and for scale points 3 and 4.

Thirty-three interviews were randomly selected for the determination of interrater reliability, 11 from each of the three interviewers. The measure of interrater reliability was the percentage of essential agreement between each interviewer and an independent rater (CB). The latter made his ratings after listening to electrical recordings of the interviews. Two ratings were said to agree if they were within one point of each other. The criterion for acceptable reliability for each variable was a percentage agreement of 80 or better.

Six of the 17 variables rated met the dual criteria of (a) satisfactory interrater reliability and (b) sufficient range to be considered worthy of further analysis.[5] The six interview variables were: (a) *expressed criticism of father,* (b) *expressed criticism of mother,* (c) *reported aggression toward father,* (d) *reported aggression toward mother,* (e) *anxiety associated with lack of achievement,* and (f) *conformity with authority.*

Modified Rorschach. The variables obtained from this instrument included: (a) *latency to phallic stimuli*—for each *S*, the latencies to the 32 stimuli were rank ordered, and the average rank for the four stimuli which were strongly suggestive of a phallus comprised this score; (b) *latency to vaginal stimuli*—four of the ink blot stimuli were suggestive of female genitals, and the latency score was the average rank for these four cards; (c) *aggression*—this score consisted of the number of responses in which people or animals were described in aggressive activity (people fighting, animals struggling), a destructive scene was perceived (landslide, A-bomb), or people, animals or objects were seen as dead or mutilated. Interrater reliability for this score was 92 per cent.

Gough Brief Femininity Scale. This scale is a 58-item self-rating test. For such items as "I want to be an important person in the community," the subject indicates whether the statement is true or false. The final score consists of the number of items where the response is coded feminine. The scoring code was derived from previous empirical studies (7). The test covers several aspects of masculinity-femininity including preferences for conventional occupational roles, sensitivity to nuances of social relationships, and concern with one's own feelings and one's own body.

Franck Drawing Completion Test.[6] This measure, in contrast with the Gough test, was designed to tap more covert levels of masculine vs. feminine identification. The Franck Test consists of 36 line drawings to be completed by the *S* in any way he wishes. The score consists of the number of items coded feminine on the basis of either the content of the drawing or its formal properties. For example, expanded drawings are scored masculine, while internally elaborated ones are scored feminine. According to Franck and Rosen (5), the test assesses unconscious levels of masculinity-femininity, levels which are hypothesized to be related to body image of the *S*. In a previous investigation (9), the coder of the present study (LL) demonstrated 90 per cent essential agreement using the Franck scoring manual (4).

---

[5] Interview variables which did not show sufficiently broad distributions were: manifest hostility during the interview, differences with father's attitudes and values, differences with mother's attitudes and values, and opposite sex interests. Variables with unacceptable reliability were: manifest dependency in the interview, affiliative needs with same sex peers, anxiety over heterosexual activity, importance of achievement, and tendency to behave dependently with others.

[6] The copyright of this test is owned by Mr. S. S. Dunn, Bureau of Educational Research, Melbourne, Australia.

French Insight Test. This test was used to assess the *S*'s preoccupation with several needs or goals. As employed in the present study, the test consisted of the 20 items used in the two forms of the original test (6). There were separate forms for boys and girls. Each item described a characteristic behavior of a boy (girl), e.g., "Joe would rather follow than lead," "John said, 'Look what I've done'," "Tom never joins clubs or social groups," and the *S* must describe the personality of the protagonist. For each item the *S* answered three questions: (a) what is the boy (girl) like? (b) what does he (she) want to have or do? (c) what are the results of his (her) behavior apt to be? The *S*s' responses to the test items were scored for preoccupation with the following needs or goals: *achievement,* when the *S*'s description of the stimulus character's behavior indicates a goal of approval (from self or others) for intellectual or vocational competence; *affiliation,* when the goal is to be liked or accepted by others or to be part of a group; *dependence,* when the goal is to obtain help or emotional support from others; *autonomy,* when the goal is to be self-sufficient, to avoid direction, manipulation, dominance, or help from others.

For each of the 20 test items, if none of the above-mentioned needs was predominant, a score of zero was given. Otherwise, the predominant need was scored. When achievement preoccupation was scored, a subscore was given for achievement-recognition if the *S* clearly stated that the major motivation for achievement was recognition from other persons.

One story at a time was rated for all *S*s before the next story was assessed, and the *S*s' identities were disguised. Interrater reliability was determined on an independent sample and exceeded 80 per cent essential agreement for each of the variables rated.

Story Completion Test. This test, a projective technique, was designed to elicit responses regarding severity of moral standards and defenses against guilt following transgression of such standards. There were separate forms for boys and girls. The test consists of three incomplete stories, each of which is about a young boy (girl) who violated an established and accepted moral standard in relationship to: (a) his parents (story 1), (b) his peers (story 2), and (c) his male employer (story 3). The *S*s were instructed to complete each story by telling how the hero feels, what happens, and how the experience turns out. This story technique and the types of code used have been described in recent publications elsewhere (1). Only the first and third stories were coded in the present study. The first incomplete story, hereafter referred to as "the boxes story" or SCB, concerned disobedience toward parents. Despite previous orders to the contrary, the central character takes some boxes down from his parents' closet. The third incomplete story, hereafter called "the photography story" or SCP, described the central character's destroying valuable photography equipment belonging to his male employer after the latter frustrates him. In both stories, the theme clearly implied that no one knew, or could find out about, the hero's (heroine's) actions.

Each story was coded for the presence or absence of several defenses against guilt. These included: (a) *denial*—denial was scored when the *S*'s response contained events which contradicted those of the story beginning; (b) *intellectualization*—this variable was scored for responses which were essentially devoid of any expression of affect; (c) *acceptance of responsibility* —this variable was scored when the hero (heroine) acknowledged responsibility for his (her) aggression and violation of a moral standard and paid for the violation in a manner that "fitted the crime."

Each story was also rated for *severity* of moral standards on a seven-point scale. This rating evaluated the degree to which the hero (heroine) punished himself, consciously or unconsciously, for his actions. One story for all *S*s was coded at a time, and all scoring was done with the *S*'s identity disguised. The coding was checked on an independent sample and found to exceed 85 per cent essential agreement for each of the four variables coded.[7]

Self-rating Inventory. This 110-item instrument was designed to assess a number of behavioral areas. Relevant areas for the present study included: (a) *aggression*, (b) *desire for acceptance*, (c) *identification with mother*, and (d) *identification with father*. The *S* was asked to place a number from 1 to 4 by each statement indicating the degree to which he felt the statement was characteristic of himself. The inventory contained 16 aggression items which included statements such as: "When I get mad at somebody, I usually tell them so"; "I get mad when someone tells me to hurry with some job." The Inventory's six acceptance items were: "I don't think you have to have friends in order to be happy"; "I'd like to be able to make friends easier"; "Occasionally I'll go along with the crowd, even though I don't want to, because I don't want to be different"; "I find it hard to say 'no' when people ask me to do something"; "I try hard to make new friends"; "I sometimes feel lonely and left out of things." The five identification with mother items were: "I agree with most of the ideas and attitudes of my mother"; "When I was a child my mother was pretty strict with me"; "I often wish my mother were smarter than she is"; "I sometimes think that I'd like to be more like my mother in temperament and personality"; "I feel that I know a little more about life than my mother." The five identification with father items were: "I agree with most of the ideas and beliefs of my father"; "I often wish my father were more successful than he is"; "I feel that I know a little more about life than my father"; "When I was a child, my father was pretty strict with me"; "I sometimes think that I'd like to be more like my father in temperament and personality." The *S*s' score for each variable was the average of his ratings for the items pertaining to that variable.

---

[7] The authors are grateful to Paula Brown, a senior in the Department of Psychology at Antioch College, for doing the reliability check.

Statistical Analysis of the Data

For each variable from each assessment technique, the distribution of scores was divided into high and low at the cutting point closest to the median score for the total sample of *S*s. This division was used for all statistical tests of sex differences. Median splits were obtained for the boys and girls separately for analyses of relations between variables. Exceptions to the above were the scores derived from the French Test and the Story Completion Test: for these tests, the terms "high" and "low" refer to the presence and absence of the variable under consideration. When more than 40 *S*s were involved in an analysis of the data, chi square corrected for continuity (11) was used; where there were fewer than 40 *S*s, the Fisher-Yates exact test was employed (3, 10). All probabilities presented in the following data summaries are for two-tailed tests.

RESULTS AND DISCUSSION

Sex Differences on Individual Aggression Variables

There were six variables concerned with aggression in the study. Four were interview variables: expressed criticism of father, expressed criticism of mother, reported aggression towards father, and reported aggression toward mother. The remaining two were aggression in the Self-rating Inventory and aggressive ink blot percepts.

Table 2 presents all the sex differences which were significant beyond the .10 level for a two-tailed test. In every case the statistic used was a cor-

TABLE 2

SEX DIFFERENCES ON INDIVIDUAL VARIABLES

Variable	*Direction*	p^*
Reported aggression toward father—I	B > G	< .05
Aggression—SRI	B > G	< .10
n Independence-autonomy—Fr	B > G	< .10
n Affiliation—Fr	G > B	< .005
Anxiety about sexuality—SRI	G > B	< .10
Gough Femininity Scale	G > B	< .005
Accept responsibility—SCB	G > B	< .005
Severity of standards—SCP	B > G	< .05

* The statistic used was chi square corrected for continuity. The probabilities are for a two-tailed test.

rected chi square. Only two of the eight involve aggression measures. Boys were higher than girls on reported aggression to father in the interview and on self-rated aggression.

Boys also scored higher than girls on independence-autonomy responses to the French Test and on severity of guilt on the Story Completion Test. In contrast, more girls than boys obtained high scores for need for affiliation as scored from the French Test and for self-ratings of anxiety about sexual activity. The girls also obtained higher scores on feminine attitudes and interest than boys and on acceptance of responsibility for destructive action in the Story Completion Test.

Most of these findings are congruent with the generally-accepted picture of sex differences in middle-class American adolescents. Adolescent boys are expected to be more aggressive than girls and to admit that they are. Boys are expected to be more preoccupied with the development of autonomy and independence while girls are more often pictured to be more concerned with affiliation and anxious about sexual activity. However, boys and girls were not significantly different from one another on four of the variables concerned with aggression: amount of expressed criticism of mother, amount of expressed criticism of father, reported aggression toward mother, and aggressive imagery on the Rorschach. As will be noted below, there were various sex differences both in the interrelationships among these aggression variables and the relationships between these aggression variables and variables assessing other behavioral areas.

Sex Differences in Relations between Aggression Variables

The relations between pairs of the six aggression variables for each sex are presented in Table 3. For these six variables, nine of the 30 exact tests reached the .10 level of significance or better for a two-tailed test. Six of these involved the boys; three, the girls.

TABLE 3

SIGNIFICANT RELATIONS AMONG AGGRESSION VARIABLES

	Expressed Criticism of Father—I		Expressed Criticism of Mother—I		Reported Aggression to Father—I	
	Boys	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*
Expressed Criticism of Mother—I	+.004					
Reported Aggression to Father—I	+.02	+.05	+.02			
Reported Aggression to Mother—I			+.01	+.02	+.02	+.03
Aggression—Rorschach			+.08			

NOTE.—The cell entries are two-tailed p values for the exact test. The plus signs indicate that all significant relations were positive.

The three significant relations for the girls involved the same pairs of variables as three of those for the boys. For both boys and girls, reported aggression to father and mother in the interview were positively related to each other. There was also a significant relation, for both sexes, between interview ratings of expressed criticism of each parent during the interview and reported aggression towards that parent.

The boys were unique in that all of their additional significant relations were associated with expressed criticism of mother during the interviews. This variable was positively related to criticism of father in the interview, reported aggression toward father in the interview, and the amount of aggression acknowledged in the Self-rating Inventory.

The expressed criticism of, and reported aggression toward, each parent, which were positively related for both boys and girls, may be two manifestations of a more general propensity toward acceptance or nonacceptance of parents during the struggle for independence typical of American middle-class adolescence. The three additional significant relations for boys may be signs of the more intense struggle for independence by adolescent males, the greater overt aggression permitted boys in middle-class society, and/or more subtle sex differences in adolescent development.

Sex Differences in Relations between Measures of Aggression and Other Variables

Of the 264 exact tests run between aggression and other variables, 23 were significant at beyond the .10 level for a two-tailed test. However, 11 of the 23 were significant at the .02 level or beyond. The data are summarized in Table 4.

Contrary to the picture for relations among aggression variables, fewer significant relations of aggression variables to nonaggressive variables were found for boys than for girls. Sixteen of the 23 were for girls; seven, for boys. The patterns for boys and girls were also dissimilar. For the boys, the focal variable was aggressive ink blot percepts; for the girls, significant relations centered around expressed criticism of, and reported aggression toward mother.

For the boys, aggressive imagery on the Rorschach was *positively* related to high feminine scores on the Franck Test and to dependency on the French Test and *negatively* related to interviewer ratings of conformity to authority. Also, reported aggression toward father was associated with self-ratings of a strong need for acceptance and with preoccupation with dependency on the French Test. It is possible that the more aggressive boys of the present sample feel free to express aggressive and dependent wishes; or, it may be that their aggression is in reaction to passive-dependent "feminine" needs.

There were two other relations between aggression and other variables for the boys. Reported aggression toward mother in the interview was positively related to achievement preoccupation on the French Test; the self-

Table 4

SIGNIFICANT RELATIONS BETWEEN AGGRESSION AND OTHER VARIABLES

Other Variables	Expressed Criticism of Father—I		Expressed Criticism of Mother—I		Reported Aggression to Father—I		Reported Aggression to Mother—I		Aggression —SRI		Aggression —R	
	Boys	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*
Achievement												
Anxiety about lack of Achievement—I	..	..	..	−.004	..	..	..	..	..	..	..	..
n Achievement—Fr	..	..	..	+.02	..	..	+.04	+.08	..	..	..	..
n Achievement-recognition—Fr	..	..	..	+.02	..	..	..	+.08	..	..	..	..
Affiliation												
n Affiliation—Fr	..	..	..	..	..	..	..	..	..	..	..	−.05
Acceptance—SRI	..	..	..	−.02	+.002	..	..	−.008	..	+.02	..	..
Dependence												
n Dependence—Fr	..	..	..	..	+.10	..	..	..	..	..	+.02	..
Conformity to authority—I	..	−.02	..	..	..	..	..	..	..	..	−.003	..
Identification												
Identification with mother—SRI	..	−.08	..	−.06	..	..	..	..	..	..	..	..
Franck Test	..	..	..	..	..	..	..	..	..	..	+.08	..
Sex												
Latency to female cards—R	..	..	..	−.05	..	..	..	..	..	..	..	..
Severity of Standards												
Severity—SCB	..	..	..	..	..	..	..	..	..	+.07	..	..
Severity—SCP	..	..	..	−.05	..	..	..	..	−.02	..	..	..
Defenses vs. Guilt												
Intellectualization—SCB	..	..	..	+.07	..	..	..	..	..	..	..	..

NOTE.—The cell entries are two-tailed test p values for the exact test. The plus and minus signs indicate the direction of relationship.

rating of aggression was negatively related to severity of standards on the photography story of the Story Completion Test.

For the girls, eight of the 17 significant results presented in Table 4 concerned relations of other variables with the girls' expressed criticism of mother during the interview. Girls who expressed relatively strong criticism toward their mothers were rated low in anxiety about achievement in the interview. They were slow to respond to Rorschach stimuli resembling female genitalia and rated themselves as being dissimilar to their mothers and having little interest in acceptance from others. In the Story Completion Test they expressed little guilt over aggressive behavior toward an authority figure (SCP); they also expressed little affect in the story about the violation of parental rules (SCB). Girls who were critical of their mothers were also preoccupied with achievement and recognition for achievement on the French Test. (For the girls only, achievement preoccupation and achievement recognition themes on the French Test were positively related at the .02 level.[8])

Three of the other significant relations for girls concerned reported aggression towards mother in the interview. Reported aggression toward mother was positively related to achievement preoccupation and to achievement recognition on the French Test and negatively related to need-acceptance on the Self-rating Inventory. These findings were parallel to those obtained for expressed criticism of mother in the interview.

There were five additional significant relations between aggression variables and other behaviors for the girls. Two of them involved expressed criticism of father in the interview. Girls who were critical of their fathers were rated low in conformity to authority by the interviewers and obtained low scores on identification with mother on the Self-rating Inventory.

Two of the three remaining significant results for the girls concerned the aggression score derived from the Self-rating Inventory. Girls who rated themselves as aggressive on the Inventory evidenced strong moral standards regarding disobedience to parental orders in the Story Completion Test, and girls who rated themselves high on aggression also rated themselves high on need for acceptance. Finally, a negative relation between the girls' Rorschach aggression score and preoccupation with affiliation on the French Test was obtained.

In summary, the results present an interesting pattern. For individual variables, there are fewer sex differences in aggression than predicted. For relations among aggression variables, there are more significant results for boys than girls, a finding which confirms the expectation that boys are more aggressive than girls during adolescence. However, the associations between the aggression variables and other behaviors give a different picture. There are more such significant associations for girls than for boys, and the patterns for the two sexes are strikingly different. Only two variables, reported aggression toward mother in the interview and achievement preoccupation on the French Test, are significantly related for both boys and girls. While for boys most of the correlations involved aggressive imagery on the Rorschach, the major emphasis for girls was on criticism of and aggression towards mother. It was suggested that the aggressive boys feel free to express various motivational content on a variety of instruments or that their aggression is in reaction to unacceptable dependent needs. For the girls, the aggressive responses towards mother were related to "masculine" interests, behaviors, and needs.

---

[8] This result does not appear in the tables of this paper.

While the meanings of the results obtained are not clear, they are in accord with Douvan's suggestion (2), based on her study of a national sample of Boy and Girl Scouts, that sex differences in American adolescents may not be so much a matter of differences in individual variables per se as in the specific types of conflicts and methods of resolution unique to each sex. In the present study the patterns of personality characteristics associated with any one aggressive variable are different for boys and girls. These differences most likely reflect the many meanings of the variables we chose to call aggressive and the fact that aggressive behavior has different meanings, outlets, motives, and defenses associated with it for boys and girls.

Summary

This report presented data from a study of normal middle-class adolescents. The results reported focused on sex differences in relations between aggressive behavior on the one hand and achievement, dependence, independence, affiliation, sex anxiety, identification, and moral standards on the other.

The subjects were 32 boys and 22 girls enrolled with their families in a longitudinal study at the Fels Research Institute for the Study of Human Development. These data were obtained from individual interviews and from scores derived from several individual and group-administered assessment techniques. These techniques included a modified Rorschach, the Gough Brief Femininity Scale, the Franck Drawing Completion Test, the French Test, a specially constructed story completion test, and a self-rating inventory.

Major results were as follows: (a) On individual variables, boys scored higher than girls with respect to reported aggression towards father in the interview, aggression on the Self-rating Inventory, independence-autonomy on the French Test, and severity of moral standards on a story completion item. The girls scored higher than the boys with respect to preoccupation with affiliation on the French Test, anxiety about sexual activity expressed in the Self-rating Inventory, femininity on the Gough Test of feminine attitudes and interests, and in taking responsibility for aggressive action toward a frustrating authority as assessed from the Story Completion Test.

(b) There were more significant interrelations among aggression variables for the adolescent boys than for the girls. Significant findings for girls were limited to relations between interview ratings of expressed criticism of, and reported aggression towards, parents. For the boys, the aggression variables of the interview related to one another as well as to aggressive Rorschach imagery. (c) There were more significant relations for girls than for boys between their aggression scores, on the one hand, and measures of the other behaviors, on the other. For the boys, both aggression towards father and Rorschach aggression were related to self-rated need for acceptance and to dependency reflected in their projective test responses. For the girls, expressed criticism of mother was related to preoccupation with recognition for achievement, anxiety about sexuality, low concern with affiliation, low desire to be similar to mother, and minimal guilt about aggressive acts directed against a male authority. Expressed criticism of father by the girls was negatively related to desire to be similar to mother and conformity to authorities. In addition, for these girls, aggressive Rorschach imagery was negatively related to preoccupation with affiliation in the French Test responses.

Implications of the results with respect to the different sex role demands and conflicts of middle-class adolescent boys and girls were discussed.

REFERENCES

1. ALLINSMITH, W. The learning of moral standards. In D. R. Miller & G. E. Swanson, *Inner conflict and defense.* Holt, 1960. Pp. 141-176.
2. DOUVAN, E. Character processes in adolescence. Paper read at Amer. Psychol. Ass., New York, September, 1957.
3. FINNEY, D. J. The Fisher-Yates test of significance in 2 × 2 contingency tables. *Biometrika,* 1948, 35, 145-156.
4. FRANCK, K. Franck Drawing Completion Test and Manual. Unpublished ms., Australian Council for Educational Research, Melbourne.
5. FRANCK, K., & ROSEN, E. A projective test of masculinity-femininity. *J. consult. Psychol.,* 1949, 13, 247-256.
6. FRENCH, E. G. Development of a measure of complex motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society: a method of assessment and study.* Van Nostrand, 1958.
7. GOUGH, H. G. Identifying psychological femininity. *Educ. psychol. Measmt,* 1952, 12, 427-439.
8. KAGAN, J. The long term stability of selected Rorschach responses. *J. consult. Psychol.,* in press.

9. Lansky, L. M. Sex identity and defenses against aggression. In D. R. Miller & G. E. Swanson, *Inner conflict and defense*. Holt, 1960. Pp. 272-288.

10. Latscha, R. Tests of significance in a 2 × 2 contingency table: extension of Finney's table. *Biometrika*, 1953, 40, 74-86.

11. Siegel, S. *Nonparametric statistics for the behavioral sciences*. McGraw-Hill, 1956.

EXPERIMENTAL DEVELOPMENT OF "LIKE" AND "DISLIKE" OF OTHERS AMONG ADOLESCENT GIRLS

EVAN R. KEISLAR*
University of California, Los Angeles

Several studies have indicated that adolescent girls who are well liked by their peers are rated as "friendly" (1, 2, 4). But such descriptive studies do not in themselves demonstrate that girls like one of their peers because she is friendly. Assuming that friendliness means the expression of social approval or "like," to establish this cause and effect relationship we must somehow contrive to have girls liked by others and then note the effect upon the girls' reciprocal liking for these others.

This problem of the exchange of social reinforcements between persons has been studied by Sidowski, Wyckoff, and Tabory (5). In a minimal social situation, two subjects in separate rooms alternately gave each other either a score or a shock, a positive or a negative reinforcement respectively, by pressing one of two buttons. Even though neither subject was told which button resulted in a score for the other, the subjects learned to give each other a high proportion of positive reinforcements.

However, the problem posed by the present investigation involves a situation where the prior learning of the subject is clearly more relevant than in the study just cited. Here the task is to find out whether adolescent girls have already learned to give positive reinforcements, i.e., expressions of "like," to those who have previously shown a liking for them. The general design involved showing subjects that they were liked by one girl and disliked by another. The effect upon their expression of "like" to each of these other two girls was then noted.

---

* Department of Education, University of California, Los Angeles 24.

Child Develpm., 1961, 32, 59-66.

One could, of course, view this expression of "like" as the result of a change in attitude toward others or a shift in a preference for others, a change which occurred when the subjects were told that they were liked or disliked. But, in this study at least, the existence of such an attitude or preference can be inferred only on the basis of the subject's behavior, i.e., his expression of "like" or "dislike." One cannot use these terms, an attitude or a preference, to "explain" this behavior. To avoid such circularity of reasoning, the expression of "like" or "dislike" is here considered simply as behavior which the subjects have previously learned to exhibit under such circumstances.

This prior learning of the subjects may be viewed more profitably within the general theoretical framework presented by Skinner (6) on verbal behavior. The expressions of "like" or "dislike" may be regarded as classes of verbal behavior of a speaker. If a speaker's verbal responses of "like" are reinforced by a listener with approval, an approving listener comes to control such expressions of "like." In other words, individuals will more often show their liking of persons who have previously approved of them.

On the other hand, if the listener does not reinforce the speaker's expression of "like" or if the listener shows disapproval, indifferent or disapproving audiences less often arouse in the speaker such verbal responses of "like." In fact, if the speaker has on some occasions shown some form of "dislike" toward a disapproving listener, e.g., "go away" or "I don't like you," and this has resulted in the removal of this source of negative reinforcement, the speaker will learn to show his dislike of persons who disapprove of him. (Other types of verbal behavior such as submissiveness or attempts to "smooth the feelings" of listeners may, of course, be learned where this kind of behavior has resulted in the reduction of the listener's disapproval.) If we assume that the processes described are the dominant ones in the subjects' past history, this line of reasoning suggests that the "approving-disapproving" characteristic of an audience partly determines the extent to which a speaker shows "like." It was the hypothesis of the present study that the subjects, viewed here as speakers, would express a greater amount of "like" toward a girl, a listener, who had previously liked them than toward a girl, another listener, who had previously disliked them.

In this investigation three adolescent girls were tested at a time in a laboratory situation where the identity of each girl was hidden from the others to eliminate the extraneous effects of appearance, prior acquaintanceship, or irrelevant behavior. The girls answered questions on their preferences and expressed "like" or "dislike" of each other by pressing buttons which, they were informed, turned on appropriate lights the others could see. Conversely, each girl saw lights which presumably revealed the responses, answers to preference questions and expressions of "like" and "dislike," of the others.

In actuality there was no communication whatsoever among the girls. The lights each girl saw were really operated by the experimenter according to a predetermined schedule. Each girl's responses were recorded but never were revealed to the others. In this way it was possible for the experimenter to control the administration of "like" and "dislike" given to each girl and hence ascertain the effect upon her own expression of "like" for the others.

Since the experimenter controlled all the information given, the three girls tested at one time were regarded as three separate subjects and received identical treatments. During the first part of the experimental session, each subject responded to a set of preference items. She was informed by means of lights, which the experimenter operated, that one of the other "girls" liked her on the basis of the way she answered the items and that the other "girl" disliked her. Following this treatment, each subject then expressed either "like" or "dislike" on the same basis to first one and then the other of the two "girls."

HYPOTHESIS

Adolescent girls express a greater liking for a girl who has previously shown a liking for them than for a girl who has previously shown a disliking for them, such liking being defined as the number of times an individual indicates "like" of the other instead of "dislike" on the basis of the way the other answered a preference test item.

METHOD

Subjects

There were 72 ninth-grade girls with Otis scores over 100 enrolled in a suburban junior high school. The subjects were tested three at a time making 24 sessions in all.

Apparatus

Three booths, each four feet square, were placed side by side with open fronts so that each subject could see items which were projected through the center booth on the screen in front. In each booth in front of the subject was a small table on which were placed four button-type switches. Two of these buttons were used to indicate answers to the two-choice items projected; the other two buttons were used to indicate "like" or "dislike." A panel of four lights on the table revealed to each subject what button she had just pressed. The numbers "1" and "2" were illuminated in this panel for item answers; a green light and a red light in the panel were turned on for "like" and "dislike," respectively. The button-pressing responses of each subject closed relays from which this information was recorded by an IBM key punch after each item.

At each side of each booth were four more lights, the numbers "1" and "2" and a green and a red light, corresponding to the panel on the table. Subjects were informed that the four lights on the left of the booth were turned on by the girl on her left (or at the other end of the row); similarly, the four lights at the right presumably indicated the responses of the girl on her right. All lights at the side were controlled, through a bank of relays, by means of a prepunched card read by the key punch. In this way each of the three subjects always observed the same pattern of lights as did the others, a schedule predetermined by the experimenter.

Since each time a girl pressed a button a relay closed with an audible click in the experimental room, it was necessary to mask the fact that the three girls were all answering at the same time. To accomplish this, a special relay was made to close and open irregularly throughout the experimental session, thus providing a continuous random sequence of audible clicks.

Test Items

All subjects were shown on the screen three series of two-choice test items adapted from Edwards' Personal Preference Schedule (3). Each series consisted of 26 items an example of which is given below:

1. I like to make new friends.
2. I like to be successful in whatever I undertake.

When responding to these items, subjects were instructed to select the statement which more nearly applied to them and to press the corresponding button.

Two forms of this test were constructed, each form consisting of 26 items. For half the subjects form A was used for the first series of items and form B for both the second and third series. The other half of the subjects were shown form B for the first series and form A for the second and third. For all subjects, therefore, the sets of items presented during series 2 and 3 were identical.

Procedure

Three girls were called out separately from different classrooms and were seated, one in each booth, without their seeing each other. Subjects were instructed to remain quiet and to ask no questions except in writing or by sign language. Even after the experimental session was over, the girls were sent back separately to their classrooms without learning each other's identity. This was done to prevent any possible embarrassment and to keep the girls from comparing notes.

Following an orientation set of six items, during which the subjects learned the purported functions of the buttons and the lights, the three girls were instructed as follows:

Three series of preference items, for which there were no right or wrong answers, would be projected on the screen in front. Each of the three girls would in turn give her answers first for one series of items. For each item, the answer given by the girl responding first would be shown by means of lights to the other two girls. The number "1" or "2" would be lighted on the side of the booth. The other two girls would then indicate, by pressing either the "like" or the "dislike" button, whether they liked or disliked the first girl "on the basis of the way she answered the item." Subjects were instructed that this judgment of like or dislike was not to be made simply because they agreed or disagreed with the girl's answer; it was pointed out that a person might like someone even though she would not answer in the same way, and, conversely, she might dislike someone who gave the same answer as herself. When the two girls had made their judgment of "like" or "dislike," the first girl would be informed of the decisions of the others by means of the red and green lights on each side of the booth. After a few seconds all lights would be turned out, a new item would appear, and the process would be repeated.

When the first series of items was completed, another girl would answer first and would be informed of the "likes" and "dislikes" of the others. For the last series of items the third girl would answer first. So that each girl could find out for which series of items she was to answer first, subjects were instructed to look at a card on their tables where this information was presented.

In actuality, since these three order cards were identical, all three girls followed exactly the same order. During the first series of items, each subject gave her responses first to the preference items. (A yellow light was flashed in her booth if she took more than 10 seconds to respond.) She was then shown, after a five-second interval, a red light or a green light, one on each side of the booth. These lights, controlled by a prepunched card, informed each subject that one girl liked her on 23 of the 26 items and that the other girl disliked her on 23 of the 26 items.

The three subjects on the remaining two series of items indicated their "like" or "dislike" of each of the "other two girls" in turn. During the second series of items each subject was shown lights on one side of the booth, purportedly the responses of the girl on that side (or at the other end of the row) to the preference test items. During the third series of items subjects were shown lights on the other side of the booth. The sequence of the illuminated numbers for series 2 was made identical with that in series 3 by using the same punched card for both series. Since the same 26 items were shown for the second and third series, as has been previously explained, and since the pattern of lights shown on one side of the booth was identical with that shown on the other, the presumed preferences of the "other two girls" were therefore identical. It was assumed consequently that any difference in the number of "likes" given by each subject to the "other two girls" would be attributable entirely to the difference in treatments received from "them" in the first series.

For half of the subjects, i.e., for 12 of the 24 groups of three subjects each, the "girl" who liked them was on the left (at least "her" answers and expressions of "like" and "dislike" appeared on the left) while the "girl" who disliked them was on the right; the opposite was true for the other half. For half of each of these sets of groups the subjects were shown the answers of the "girl" who liked them on series 2 and those of the "girl" who disliked them on series 3; for the other half of each of these sets of groups the subjects saw the answers of the "girl" who disliked them on series 2 and those of the "girl" who liked them on series 3. Finally, for each of these subsets, half of the groups of subjects were shown form B of the preference test for series 2 and 3; for the other half, form A was used for these last two series.

RESULTS

The mean number of "likes" given to the girl who liked the subject previously was 19.1 with a standard deviation of 2.8. The comparable figure for the girl who disliked her was 17.7 with a standard deviation of 2.3. The correlation between the two measures, based on 72 cases, was .53. To test the significance of this difference an analysis of variance was carried out analyzing the total variance into "within" and "between" subjects. The factors were the two tests of number of "likes" obtained from the last two series (T), two positions, right or left of the booth for display of the responses of the girl who liked them (P), two orders of presenting the other

TABLE 1

ANALYSIS OF VARIANCE OF NUMBER OF "LIKES" GIVEN TO TWO GIRLS, ONE APPROVING, THE OTHER DISAPPROVING

(Given by 72 adolescent girls)

Source	*df*	*Mean Square*	*F*
Between *Ss*	71		
P (Position)	1	6.6	...
O (Order)	1	25.8	...
P × O	1	73.8	5.8*
Error (b)	68	12.8	...
Within *Ss*	72		
T (Tests)	1	70.8	16.1**
T × P	1	.1	...
T × O	1	.6	...
T × P × O	1	6.6	...
Error (w)	68	4.4	
Total	143		

* Significant at .05 level.
** Significant at .001 level.

girls' responses (O), with 18 subjects in each subgroup. The effect of the different forms was not analyzed. Since nine subjects in each subgroup were shown form A for the last two series and the other nine in each subgroup form B, this factor was adequately counterbalanced. Results of this mixed-design analysis are presented in Table I. The *F* for the two tests is significant at the .001 level; all other "within" subjects effects are not significant. The hypothesis of the study is therefore clearly supported.

Conclusion and Discussion

Adolescent girls, comparable to those used in this sample, express a greater amount of liking for other girls who have previously shown considerable liking of them for their preferences than they express for girls who have previously shown considerable disliking of them on this same basis. This finding, however, must be viewed as something more than a simple exchange of social reinforcements. It was hypothesized in this study that this expression of "like" is verbal behavior which previously has been conditioned to occur in the presence of an approving audience. A test of such an hypothesis will require the use of much younger children as subjects.

Furthermore, it is probable that the subject's expression of "like" was controlled by intraverbal associations, or, to put it another way, was a result of mediated generalization. For example, during the first series most subjects may have made, in spite of the experimenter's effort to discourage them by his instructions, certain covert verbal responses, e.g., "The girl on my left has my preferences." This verbal response may have controlled the expression of "like" if, in the past, subjects have been more frequently reinforced for expressing "like" to listeners whose preferences were similar to their own. Other intraverbal associations may have been acquired by various methods more fully discussed by Skinner (6).

The use of measures of central tendency in this investigation probably has hidden reliable individual differences among these girls. For example, many girls gave more "likes" to the girl who previously disliked them than to the girl who previously liked them. Although in many cases this was probably a matter of chance, in some it may have represented a reliable pattern of behavior; these latter subjects may have "tried to be extra nice" to the disapproving girl. As was suggested parenthetically in the introduction, some persons may have learned to express "like" to a disapproving listener since this has frequently resulted in reducing the disapproval. The reinforcements for such behavior may have also come from "polite" parents.

It has been noted that the present findings are probably a reflection of the prior reinforcement histories of these subjects. Since reinforcements for such verbal behavior as "like" and "dislike" are probably a function of the culture from which the subjects come, it is pertinent to note that the subjects were all above average in intelligence and came from a middle-class suburban area. If it is true that their parents and peer groups place greater importance upon "being nice to everyone" than do individuals from other social classes, with a different sample the effect observed in this study might be considerably larger.

REFERENCES

1. ANASTASI, A., & MILLER, S. Adolescent "prestige factors" in relation to scholastic and socioeconomic variables. *J. soc. Psychol.*, 1949, 29, 43-50.
2. AUSTIN, M. C., & THOMPSON, G. C. Children's friendships: a study of the bases on which children select and reject their best friends. *J. educ. Psychol.*, 1948, 39, 101-116.
3. EDWARDS, A. *Personal preference schedule.* Psychological Corporation, 1954.
4. KUHLEN, R. G., & LEE, B. Personality characteristics and social acceptability in adolescence. *J. educ. Psychol.*, 1943, 34, 321-340.
5. SIDOWSKI, J. B., WYCKOFF, L., & TABORY, L. The influence of reinforcement and punishment in a minimal social situation. *J. abnorm. soc. Psychol.*, 1956, 52, 115-119.
6. SKINNER, B. F. *Verbal behavior.* Appleton-Century-Crofts, 1957.

EFFECT OF WITHDRAWAL OF POSITIVE REINFORCEMENT ON AN EXTINGUISHING RESPONSE IN YOUNG CHILDREN[1]

DONALD M. BAER*
University of Washington

Punishment may be defined systematically in either of two ways: (a) as the presentation of a negative reinforcer for a response or (b) as the removal of a positive reinforcer for a response (9). The first of these methods has seen extensive study in lower organisms (e.g., 4, 8). The second has been subject to a certain amount of discussion in psychoanalytic terminology ("loss of love"), but has seen very little experimental investigation in any species. There are only three exceptions apparent to this statement. Ferster (5) has used a withdrawal procedure ("time out") in studies with lower animals; Brackbill and O'Hare (3) have used it as a punishment technique in facilitating discrimination in young children; and Lewis (6) has used it to set up schedules of reinforcement in young children.

Yet the withdrawal of positive reinforcement as a punishing technique could prove a very significant process in the reinforcement history of the child. Punishment through the presentation of negative reinforcement is typically claimed to be ineffective in weakening responses. Estes concluded that ". . . a response cannot be eliminated from an organism's repertoire more rapidly with the aid of punishment than without it" (4, p. 37). Sears *et al.* claim that punitiveness in a parent ". . . is ineffectual over the long term as a technique for eliminating the kind of behavior toward which it is directed" (7, p. 484). But they also point out that half of the parents they interviewed use punishment at least moderately often and claim that spanking does some good. In the case of the child, it will often prove impossible to present a negative reinforcer without simultaneously withdrawing positive reinforcers. The parent who slaps or spanks almost inevitably will withdraw approval and the more obvious tokens of affection at the same time. Perhaps it is this withdrawal of reinforcement which is the significant technique here.

---

* Gatzert Institute of Child Development, Department of Psychology, University of Washington, Seattle 5.

[1] The research reported herein was supported in whole by Public Health Service grant M-2208 (C1), United States Department of Health, Education, and Welfare, to whom the author is indebted. Three other debts must be acknowledged: to Mrs. Josephine Reed, director of the Seattle Day Nursery Association Schools, for gracious and generous cooperation; to Mrs. Anne Pilisdorf for her reliable and intelligent performance as *A*; and to Dr. Sidney Bijou for his continuing support, encouragement, and advice.

Child Develpm., 1961, 32, 67-74.

In any case, experimental studies of the punishing effect of the withdrawal of positive reinforcement in the child should prove of systematic and practical value. The present study was designed to yield data on the effectiveness of reinforcement withdrawal for the sole purpose of weakening a response in a situation where a competitive response is not deliberately strengthened through positive reinforcement.

METHOD

Subjects

Children from a day care center of the Seattle Day Nursery Association were used. All *S*s were experimentally naive. At the outset of the experiment, 16 *S*s were divided randomly into two groups, Punishment and Control. During the course of the experiment, which required that each child participate on five different sessions, six *S*s were lost, usually to their family vacation, occasionally to illness, and in one case, to withdrawal of the child from the school. There remained 10 *S*s, five in each group. The average age of the Punishment group was 6-10, and of the Control group, 6-6. All *S*s were within the normal range of intelligence and represented lower income groups.

Apparatus

The entire experiment was conducted in a mobile laboratory built into a 19-foot house trailer (2), which was parked inside the play yard of the school. The interior of the laboratory included a one-way observation booth for the experimenter and a playroom for the child. The playroom contained a small chair, two tables holding toys or apparatus, a screen on which could be projected cartoons, and a partitioned corner in which an accompanying adult (*A*) could sit, out of the child's sight but still present.

The experiment centered about three pieces of apparatus. The first of these was a translucent screen, on which the experimenter could project cartoons. The screen was mounted in the wall separating child and experimenter; the movie projecter, a Busch "Cinesalesman," operated from the experimenter's side. The projector contained three cartoons on an "endless" reel of film and could present these in an uninterrupted sequence for an indefinite number of cycles. The cartoons were of the Castle Films Woody Woodpecker series (Woody Plays Santa Claus, The Hollywood Matador, and The Dizzy Acrobats); each was black and white, with sound, and lasted seven minutes.

A second piece of apparatus was a white box, approximately one foot on a side, mounted on one of the tables such that *S* could view cartoons and play with this "toy" at the same time. A bar projected from the front surface of the box. At the lower right side of the front surface was a slot from which peanuts could roll as a reinforcement for pressing the bar. Every bar press resulted in a mild buzz which lasted one second, but did not necessarily produce a peanut. A bar press could also result in a two-second interruption of any cartoon being viewed at the time, if the experimenter chose.

The third piece of apparatus was a commercial toy (Playskool) which was mounted on the other table. It consisted of a cage of sticks, inside of which a small ball could be made to jump up and down a devious route by flipping a small lever at the base of the toy. This toy was not within reach of *S* as he sat viewing cartoon.

Procedure

*S*s were dealt with entirely by a young female adult, *A*. Before the experiment started, *A* had been a constant member of the play group and was thoroughly familiar with the *S*s. Furthermore, she had told them that a trailer was coming and that all the *S*s would be allowed to enter and play games. This generated a great deal of enthusiasm.

Each *S* was brought to the laboratory on five separate occasions by *A*. The first of these was to adapt the *S* to the laboratory and acquaint him with the cartoons. *A* would seat the *S* in the correct place and retire to her corner. Then the cartoons would start, and all three would play without interruption. At the end of the third cartoon (i.e., after 21 minutes), the projector was stopped, and *A* came out of her corner and took *S* back to the play yard.

The second visit was to establish a learned response in each *S*. Prior to this visit, *A* would ask *S* if he would like to do something different that day. When *S* entered the laboratory, *A* then showed him the toy which could dispense peanuts, saying, "This is a peanut machine." The *S* was seated by it (the same seat as visit 1), and *A* retired to her corner. In very short order, *S* would press the bar and receive a peanut. This started a series of bar presses, of which numbers 1, 2, 3, 5, 7, 10, 15, and 20 were reinforced with a peanut. Immediately after the twentieth response, as *S* was eating the peanut, *A* came out of her corner announcing, "That's all for today." They then left together before the child could respond further.

The third visit was to begin extinction of the learned response, punishing it as well in half of the *S*s. Prior to this visit, *A* asked *S* simply if he would like to come again. Inside the laboratory, *A* seated *S* in the same seat and retired to her corner. The cartoons were started immediately. *S* was free to watch cartoons and to press the bar as he chose. For the five *S*s of the Punishment group, any bar press during this session would turn off the ongoing cartoons (picture and sound) for two seconds. For the five Control *S*s, the bar press had no such action. In either case, the bar press produced the usual buzz, but no peanuts at any time. At the end of the third cartoon the projector was turned off, but *A* remained in her corner for at least two minutes, saying nothing. This gave *S* an interval during which the bar might be pressed, and there were clearly no cartoons to be turned off. Almost every *S* requested to leave by two minutes after the end of the cartoons. In the case of early requests, *A* was instructed to delay until the two minutes had passed to give each child equal opportunity with the bar. Thus, the visit lasted 23 minutes.

The fourth visit was to allow *S* to show any spontaneous recovery of the learned response, without punishment. Prior to this visit, *A* asked the child if he would like to see the cartoons again. Inside the laboratory, *S* was seated in the usual seat, *A* retired to her corner, and the cartoon was started. *S* could press the bar if he chose, but on this occasion the bar press did not turn off the cartoon, nor did it result in any peanuts. It did produce a buzz as before. Because it was expected that spontaneous recovery would be slight in either group only one cartoon was shown. Again, *S* was kept for two minutes after the end of the cartoon. Thus, this visit lasted nine minutes.

The fifth visit was to examine the strength of the learned response in a situation where punishment was impossible. Prior to this visit, *S* was asked simply if he wanted to come again. Inside, he was told that there would be no cartoons today, but that he could play with anything he wanted. *A* then retired, leaving *S* with only two objects capable of play: the peanut apparatus and the Playskool toy. The peanut machine buzzed but did nothing else when the bar was pressed. This session lasted five minutes, which is rather long for children thoroughly adapted to the laboratory, with only two simple toys to play with.

In summary, then, *S*s were successively adapted to the laboratory and the cartoons, taught a bar pressing response, allowed to extinguish this response with or without punishment, and allowed to show any spontaneous recovery of this response, in situations where the means of punishment was present and absent. This sequence is summarized in Table 1. The interval between the first and second visits was seven days (plus or minus two);

TABLE 1

SEQUENCE OF EXPERIMENTAL CONDITIONS FOR PUNISHMENT AND CONTROL SUBJECTS

Visit	*Purpose*	*Number of Cartoons*	CONSEQUENCE OF BAR PRESS *Punishment Group*	 *Control Group*
1	Adaptation	3	bar not present	bar not present
2	Strengthen bar press	none	peanuts	peanuts
3	Extinction —session 1	3	cartoon interruption	none
4	Extinction —session 2	1	none	none
5	Extinction —session 3	none	none	none

between the second and third visit, three days (plus or minus one); between the third and fourth visit, five days (plus or minus one); and between the fourth and fifth visit, three days (plus or minus one). All experimentation was conducted during late afternoon hours.

RESULTS

Recordings were made of the bar pressing response on a Gerbrands Harvard Cumulative recorder. Photographic facsimiles of these response curves are shown in Figure 1. Session 1 in this figure refers to *S*'s third visit to the laboratory, during which the bar pressing response resulted in an interruption of the cartoon for Punishment *S*s and otherwise represents an extinction period for all *S*s. Session 2 refers to S's fourth visit, during which one cartoon was seen without punishment. Session 3 refers to *S*'s last visit, during which he played freely with either the bar or the Playskool toy and no cartoons were shown.

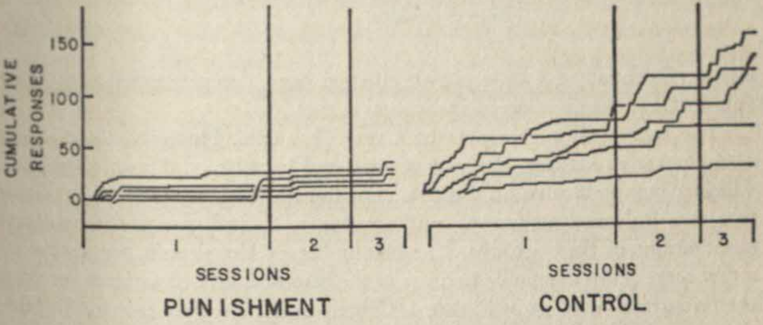

FIGURE 1—Individual cumulative response curves of Punishment and Control *S*s during repeated sessions.

It is apparent from Figure 1 that the interruption of the cartoon serves to reduce sharply the number of extinction responses *S*s make to the bar. Furthermore, this reduction seems reasonably durable. During the last two minutes of session 1, Punishment *S*s show no particular tendency to increase responding, although there are no cartoons present, and hence they cannot be punished by their interruption. During session 2, where Control *S*s show considerable spontaneous recovery, Punishment *S*s make either one or two responses (which are unpunished) and then cease responding, nor do they respond during the two minutes after the cartoon ends. During session 3, Punishment *S*s direct their play to the Playskool toy rather than to the bar far more than do Control *S*s.

These individual response curves are summarized in Figure 2, which shows the mean number of bar presses made by each group during each

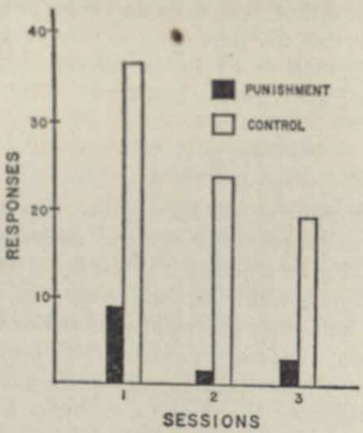

FIGURE 2—Mean responses of Punishment and Control *S*s during repeated sessions.

session. Essentially the same picture emerges from a consideration of grouped data as from examination of individual curves.

The individual curves present a very clear case. However, there is one case of overlap between the two groups, and because of this an analysis of variance was performed on the data. The Punishment and Control *S*s constitute two independent groups, and sessions 1, 2, and 3 represent repeated observations of these groups. Use of number of bar presses per session as a raw score is not justifiable because of the heterogeneity of variance between the two groups evident in Figure 1. Consequently, a square root transformation was applied. The transformed data, subjected to Bartlett's test, yielded

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF RESPONSES DURING REPEATED SESSIONS

Source	*df*	*Mean Square*	*F*	*p*
Subjects	9			
Punishment	1	71.10	11.83	.01
Error	8	6.01		
Within Subjects	20			
Sessions	2	11.00	10.78	.01
Sessions × Punishment	2	1.24		
Error	16	1.02		
Total	29			

a χ^2 of 5.28 ($df = 5, .30 < p < .50$). A summary of the analysis of variance of the transformed data is given in Table 2. Punishment and Control *Ss* are clearly distinct at the .01 level. The Sessions main effect is of no great consequence, since response is bound to decrease during repeated extinction sessions, especially when the sessions are of successively shorter duration. The absence of any interaction between the Punishment groups and Sessions makes for a gratifyingly simple conclusion.

None of the Punishment *Ss* showed any obvious signs of emotionality during punishment. One *S* in this group, after making several (punished) responses during session 1, was seen to very carefully take hold of the bar, slowly press it down, and, as the cartoon went off, whip his hand away from the bar, saying, "Wow!" Apart from this, all *Ss* took the experiment with great equanimity.

DISCUSSION

From the results obtained, it seems clear that the interruption of a cartoon which the child is viewing will serve very effectively as a punishing event: it will efficiently depress a recently strengthened response in the child's repertoire. The thesis that withdrawal of positive reinforcement is punishing is thereby strengthened, specifically for the case of the child. And it is tempting to wonder if the withdrawal of positive reinforcers may not possess punishing characteristics qualitatively different from the presentation of negative reinforcers.

An essential assumption here is that the cartoons are indeed positively reinforcing stimulation. Other research by the author with the same three cartoons indicated that they have a high and very uniform reinforcing value for children in this age range (1). Testimony to this assumption is that almost any child will learn quickly to press a bar to turn the cartoons back on when they are turned off by the experimenter every few seconds.

It may be important to stress certain aspects of this experiment which could prove critical to its conclusions. One of these is the fact that the response weakened in this procedure is a weak one of short standing. Eight peanuts cannot be taken as a significant amount of reinforcement to a healthy child (who has had juice and cookies in school the same afternoon). These reinforcements, with the increasing ratio schedule on which they are delivered, are sufficient to set up a response strong enough to be studied and manipulated, clearly, but may not give the response much resemblance to other behaviors of the child with more powerful and extensive histories of reinforcement.

Furthermore, in this study punishment has been effected in a very precise manner. Reinforcement was withdrawn immediately consequent to a response in a very consistent way. The fact that this procedure is effective may not guarantee that in a more typical situation, where punishment is offered to a child late, inconsistently, and perhaps incomprehensibly, the effect would be the same.

Finally, the effect of punishment in this study seems durable in terms of the duration of the study. But this is not a great deal of time. Possibly, given more time in the situation, Punishment *S*s would emit responses equal in number to those of Control *S*s. It seems more likely, though, that the *S*s would satiate with the entire situation before this would happen.

Apart from these qualifications to any widespread generalizations, the data give a clear picture of stimulus control in depressing a response in young children. The mechanism is best stated as the withdrawal of positive reinforcement. The technique used here would seem to be a fruitful one for further study of this problem in the laboratory.

SUMMARY

Young children were taught to press a bar for peanuts. During later extinction of this response, five Punishment *S*s were punished by turning off a cartoon they were watching at the time for two seconds as a consequence of every response. This served to depress the response considerably, relative to five Control *S*s who were not punished for responding while watching the same cartoons. During later sessions, Control *S*s showed considerable spontaneous recovery of the response, and Punishment *S*s did not, even though no longer punished. It was concluded that the withdrawal of positive reinforcement is an effective technique of punishment, at least in situations like the experimental one.

REFERENCES

1. BAER, D. M. Escape and avoidance response of preschool children to two schedules of reinforcement withdrawal. *J. exp. anal. Behav.*, 1960, 3, 155-159.
2. BIJOU, S. W. A child study laboratory on wheels. *Child Develpm.*, 1958, 29, 425-427.
3. BRACKBILL, Y., & O'HARE, J. Discrimination learning in children as a function of reward and punishment. Paper read at Western Psychol. Ass., Eugene, Oregon, 1957.
4. ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, 57, 1-40.
5. FERSTER, C. B. Control of behavior in chimpanzees and pigeons by time out from positive reinforcement. *Psychol. Monogr.*, 1958, 72, 1-38.
6. LEWIS, D. J. Partial reinforcement in a gambling situation. *J. exp. Psychol.*, 1952, 43, 447-450.
7. SEARS, R. R., MACCOBY, E. E., & LEVIN, H. *Patterns of child rearing.* Row, Peterson, 1957.
8. SKINNER, B. F. *The behavior of organisms.* Appleton-Century-Crofts, 1938.
9. SKINNER, B. F. *Science and human behavior.* Macmillan, 1953.

AN EXPLORATION OF PATTERNS OF IMPULSE CONTROL IN EARLY ADOLESCENCE[1]

NORMAN LIVSON *and* WANDA C. BRONSON*
University of California, Berkeley

Somewhere near the core of any current personality theory which lays claim to comprehensiveness there will be found some construct dealing with the fact that most adults are able to control (more or less) and to modify (in any number of ways) the direct expression of their biological impulses. In exploring the development of this phenomenon, some investigators feel more comfortable working within a psychoanalytic framework (e.g., 11); others find learning theory schema more congenial (e.g., 13) or seek to exploit the suggestions provided by a topological model (e.g., 4). We may prefer to speak of "ego development" under the constraints of the "reality principle" or of "socialization" achieved through "learned inhibitions" or of "differentiation of the internal need system" in response to forces in the "life-space." Putting theoretical and stylistic qualms aside, we can probably agree that ". . . by gaining control over the demands of the instincts, by deciding whether they shall be allowed to obtain satisfactions, (and) by postponing satisfaction to times and circumstances favorable in the external world . . ." (5, p. 15), childhood impulsivity becomes gradually transformed into adult responsibility.

While anthropological and psychoanalytic observations and theory regarding the development of impulse control have attended to the full span from infancy to adulthood, empirical investigations leave a gap in the years between childhood and late adolescence. If, as is generally maintained, early adolescence represents a transitional stage, a period of upsurgence of biological impulses, then the modes for handling tensions which the child has already developed would tend, at this time, to assume an increasingly influential role in his psychological functioning. Assuming this, the intent of this study is to explore the linkages of the ways in which early adolescent boys and girls handle their impulses with other aspects of their behavior and personality organization and, further, to examine these relations in the context of the specific values of their possibly very special "subculture." In confining this investigation to the period of early adolescence, we leave aside issues concerning both the genesis of ego control patterns and their

---

* Guidance Study, Institute of Human Development, University of California, Berkeley 4.

[1] We are pleased to acknowledge the contributions to this report of the staff, past and present, of the Guidance Study and of its director, Dr. Jean W. Macfarlane. We are particularly indebted to Erik Homburger Erikson for his case evaluations, to Dr. Betty Kalis for working out a method for quantification of this material, and to Dr. Mary Darby Rauch for help in carrying out quantification procedures. Drs. Marjorie Honzik and Jack Block have offered helpful criticism of the manuscript.

Child Develpm., 1961, 32, 75-88.

stability over time under the pressures of biological maturation and changes in psychosocial environment.

METHOD

The source data for this study are comprehensive clinical descriptions of a group of children, prepared in their early adolescence (ages 11 to 13) by E. H. Erikson, after a thorough review and evaluation of all material available from the Guidance Study at that time. The sample employed for this report consists of the 40 children (20 boys, 20 girls) for whom such summaries were completed and represents a rather informal selection from a much larger pool of Guidance Study developmental case histories. The Guidance Study itself is a longitudinal investigation of physical, mental, and personality development in a group of over 200 children who are representative of all children born in Berkeley, California in the years 1928-1929. [For a full presentation of this research program *see* (9).]

Erikson's personality descriptions were made amenable to quantitative analysis through a Q-sorting procedure (8) applied independently by two clinical psychologists to each summary, which yielded composite (summed) item placement scores. From the initial set of 102 Q-sort items, a selection was made of the most highly reliable and discriminating items. The resultant pool of 32 items, and their interrater agreement, is presented in Tables 1 and 2.

RESULTS

Bivariate plots of two Q-sort items, "is undercontrolling; acts upon impulse" and "is rigid, overcontrolling, inflexible," are presented in Figure 1 for boys and for girls.[2] The intercorrelations (*gamma* coefficient[3]) are, respectively, —.44 and —.53. Rather than combining these two measures to obtain a single score for degree of impulse control, each individual was assigned to an ego control group on the basis of his two-variable pattern. This procedure allows for the possibility that such a two-dimensional ego control pattern might provide us with an increment of reliable variance beyond that which could have been carried by some type of single composite score.

---

[2] These items are two of many which were adopted into the present Q-sort pool by Kalis (8) from an earlier version of the California Q-set (2). Although Block and his colleagues are explicit in their conceptualization of "ego control" as a single continuum [cf. (3)], they require separate assessments of degree of "overcontrol" and "undercontrol" in order to permit detection of "bimodal controllers" (individuals who exhibit characteristics of both ends of the dimension) (1). Our use of these two items is a matter of convenience and implies no allegiance to the theoretical model from which they derive; furthermore, the employment of the term "ego control" in this report is casual and intends no distinction from such synonymous (for us) terms as impulse control.

[3] The *gamma* correlation coefficient (6), a variant of *tau*, expresses the degree of covariation between two sets of scores, each of which represents measurement on at least an ordinal scale. Order comparisons of pairs of scores where one or both comparisons involve ties do not affect the value of the coefficient.

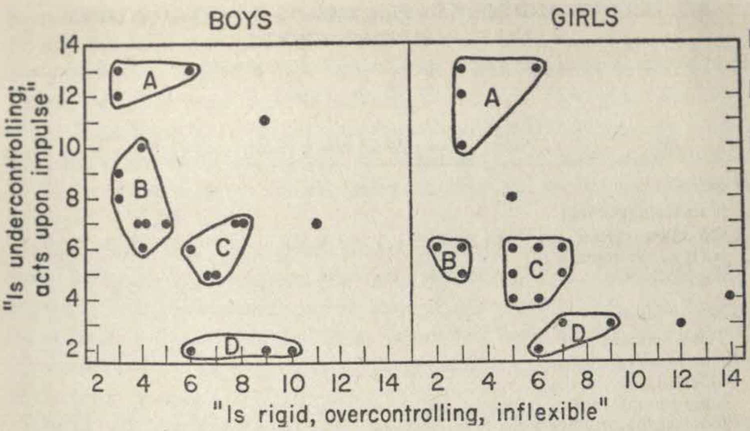

FIGURE 1—Bivariate distributions of ego control scores for boys and girls. The boundaries of the ego control groups, which were determined by visual inspection, are indicated on the scatterplots.

The groupings indicated on the scatterplots were determined by visual inspection with the requirement of at least three cases to form a group. While admittedly failing to insure an objective and unequivocal assignment of cases, this procedure is unbiased: no additional data were consulted and we had no theoretical expectations as to meaningful ego control configurations. The validity of such groupings will be indicated by their ability to generate meaningful distinctions in other aspects of personality organization and functioning.

Derivation of Cluster Scores

Since an examination of the content of the remaining 30 Q-sort items suggested that some reduction to more general dimensions might be possible, separate cluster analyses of boys' and girls' interitem correlation matrices were carried out. Trial clusters were established for each sex on the criterion that each item in a cluster shows a median *gamma* coefficient of at least .40 (approximately at the .05 level[4]) with the other items in that cluster. The results of these analyses, which permitted items to appear in more than one cluster, were examined to determine which of the trial clusters were found for both boys and girls. Only those items which showed similar patterns of interrelations for both sexes were submitted to a new cluster analysis, this time employing interitem correlations computed for the total sample. Final common clusters were established on the following criteria: (a) each item shows a median correlation of at least .50 (approximately $p = .001$) with the other cluster members and (b) this median value is higher with

---

[4] The significance level of the *gamma* coefficient can be estimated from current work on its sampling distribution carried on by Rosenthal (12).

TABLE I

RELIABILITIES AND MEDIAN PLACEMENTS OF CLUSTER ITEMS FOR EGO CONTROL GROUPS

		BOYS				GIRLS			
Item	Reliability†	A $N=3$	B $N=7$	C $N=5$	D $N=3$	A $N=4$	B $N=3$	C $N=7$	D $N=3$
Ego Control									
2. Is undercontrolling; acts upon impulse	.62	13	7	6	2	12.5	6	5	3
6. Is rigid, overcontrolling, inflexible	.57	3	4	7	9	3.0	3	6	7
Ego Strength Cluster									
5. Tends to make personally relevant many different areas; to become ego-involved	.59	6	9	5	6	7.0	9	11	7
11. Is relaxed, easy-going	.74	3	10	5	5	4.0	7	10	6
13. Absorbs family's cultural values without strain	.47	5	11	6	6	6.0	5	11	10
4.* (Does not) lack organization, synthesizing of ego strengths	.35	6	12	6	8	4.5	10	10	9
7.* (Does not) have poor controls for anxiety	.71	3	11	5	8	4.0	10	12	11
Social Adjustment Cluster									
1. Is self-confident and assured in relations with others	.70	8	11	4	3	5.5	8	10	10
16. Is accepted as a leader by peers	.62	7	8	7	6	5.0	9	9	10
17. Is popular with other children	.82	8	8	7	7	4.5	10	11	9
18. Has good role variability	.64	6	10	4	3	4.5	11	12	6
20.* Is (not) awkward, clumsy in social situations	.70	7	10	6	8	6.5	11	12	13
22.* Is (not) socially constricted; (does not have) a narrow range of social behavior	.75	9	11	5	5	7.0	12	12	8
Emotional Expansiveness Cluster									
19. Tends to be outgoing in social situations	.59	11	11	7	4	11.0	11	12	8
21.* Is (not) aloof, distant in personal interactions	.75	11	7	4	6	10.5	8	8	6
24.* Is (not) bland, emotionless in approach to situations	.61	12	10	6	5	11.5	10	11	8
32.* Is (not) emotionally constricted	.52	10	9	4	3	8.5	10	11	4

* Indicates "negative" cluster item for which scale value is reversed. Material in parentheses brings the item into proper orientation with the cluster.

† Agreement between two sorters evaluated by the *gamma* correlation coefficient (6).

its own than with any of the other clusters. Therefore, in this final clustering, no item can appear in more than a single cluster. Also, at least three items are required to define a common cluster. This procedure yielded three clusters which seemed readily identifiable as Ego Strength (five items), Social Adjustment (six items), and Emotional Expansiveness (four items). The item composition of the clusters is presented in Table 1, and a listing of the 15 residual items which failed to meet the clustering criteria may be found in Table 2.

Each child's score on a given cluster is his mean item placement over all items comprising the cluster, the scale values for "negative" items being reversed. Since item scores are the sum of placements by two sorters on a seven-category Q sort, the resultant values for both cluster and individual item scores may range from 2 to 14, higher scores denoting a greater amount of the given quality.

Intercluster (*gamma*) correlations based on these scores are as follows: Ego Strength × Social Adjustment (boys, .57; girls, .77); Ego Strength × Emotional Expansiveness (boys, .36; girls, .31); Social Adjustment × Emotional Expansiveness (boys, .46; girls, .40). For both sexes, the interrelations among clusters are quite substantial (especially when the reliabilities of the component items are considered) so that a general factor, perhaps loosely defined as "psychological health," might be substituted for the three separate clusters. However, since the items within each cluster could be sensibly assigned to conceptually independent dimensions, it seemed legitimate—and possibly useful—to keep them distinct.

Characteristics of the Ego Control Groups

In this attempt to define the correlates of the different ego control patterns we shall rely primarily upon group differences in the three cluster scores. However, a reading of the median group values for the unclustered (residual) items (Table 2) will be employed to evaluate and supplement the resultant characterization of the ego control groups. Two criteria are used in reporting discriminating residual items: (a) all extreme items (scale values above 10 or below 6) are cited as "descriptive" of a given group and (b) statements of intergroup "differences" require at least a two-point scale difference. The significance levels of intergroup differences (cited in Table 2) should be regarded as only rough guides to their relative reliability, due to the low and highly variable power of statistical evaluations with such small groups.

Boys. Figure 2 (left) summarizes the median cluster scores for the four boys' ego control groups. (The median values for individual items within each cluster are presented in Table 1.)

Group A, the most undercontrolled and the most not overcontrolled pattern, shows lowest Ego Strength, moderate Social Adjustment, and (not surprisingly in light of its definition) the most extreme degree of Emotional Expansiveness. The syndrome becomes more recognizable when we con-

TABLE 2

RELIABILITIES AND MEDIAN PLACEMENTS OF RESIDUAL ITEMS FOR EGO CONTROL GROUPS

		BOYS					GIRLS				
Item	*Reliability*†	A *N*=3	B *N*=7	C *N*=5	D *N*=3	*Diff.**	A *N*=4	B *N*=3	C *N*=7	D *N*=3	*Diff.**
3. Is independent in thoughts and feelings	.55	5	9	4	7	A<B .12 B>C .02 C<D .15	5.5	4	9	10	....
8. Shows conflict about sexual identification	.48	11	6	12	11	B<C .05	11.0	7	9	11	A>B .06
9. Identifies with mother or mother-figures	.69	9	7	10	7	C>D .06	12.0	10	12	10	....
10. Identifies with father or father-figures	.53	9	10	8	11	C<D .04	8.5	5	7	7	....
12. Represses unacceptable impulses	.52	8	9	9	9	....	6.0	12	9	10	A<B .06 A<C .03 A<D .06 B>C .03
14. Is motivated in behavior by guilt feelings	.79	8	5	8	6	B<C .15	6.5	8	7	7	....
15. Is individualistic, nonconforming	.40	8	5	5	8	A>C .08	10.0	3	4	4	A>B .06 A>C .05 A>D .09
23. Is openly aggressive, hostile in relations with others	.44	11	4	6	6	....	9.5	5	5	5	A>B .09 A>C .04 A>D .06
25. Is a verbally expressive individual	.49	8	8	6	11	....	9.5	10	10	8	....

(*table continued on next page*)

Table 2 (*continued*)

RELIABILITIES AND MEDIAN PLACEMENTS OF RESIDUAL ITEMS FOR EGO CONTROL GROUPS

		Boys					Girls				
Item	*Reliability*†	A $N=3$	B $N=7$	C $N=5$	D $N=3$	*Diff.**	A $N=4$	B $N=3$	C $N=7$	D $N=3$	*Diff.**
26. Has passive approach to environment	.73	4	4	12	9	A<C .04 A<D .10 B<C .11	2.5	4	5	3	A<C .14
27. Seeks attention and praise	.67	13	10	10	12	A>C .12	13.0	10	9	11	A>B .09 A>C .05
28. Tends to take roles of mothering as reflection of nurturant needs	.66	7	8	6	7	A>C .15 B>C .07 B>D .12	10.0	12	8	13	C<D .10
29. Is unduly fearless, counterphobic	.60	12	5	6	12	A>B .04 A>C .04 B<D .12 C<D .15	10.0	5	6	4	A>B .09 A>D .06
30. Is dominating, assertive	.43	10	6	7	7	A>C .04 A>D .10	10.5	7	11	8	....
31. Is timid, fearful	.35	6	8	11	7	A<C .04 A<D .10 B<C .10 C>D .04	5.5	10	6	8	A<B .12 B>C .12

---

* Group differences evaluated by Mann-Whitney U test ($p \leqq .15$ reported).

† Agreement between two sorters evaluated by *gamma* correlation coefficient.

sider the tendency of these boys to be fearless and counterphobic (item 29), hostile (item 23), and dominating (item 30), and not passive (item 26). However, they lack independence (item 3) and require a great deal of attention and praise (item 27). Thus, while their undeniable assertiveness may earn some degree of social success within the peer group, it appears to serve the defensive function of masking an underlying lack of security.

Group B, while also not overcontrolled but appreciably less impulsive than group A, presents a cluster profile suggestive of a uniquely high level of psychological health: these children score high, both relatively and absolutely, on Ego Strength and Social Adjustment and, while also scoring relatively high on Emotional Expansiveness, they show somewhat more restraint than the group A boys. This impression is enhanced by the residual items: of all groups, group B boys are the most independent in thoughts and feelings (item 3), show least conflict over sexual identification (item 8), and are not motivated by guilt (item 14). Further, they are neither counterphobically fearless (item 29) nor hostile or aggressive (item 23), but any suggestion of general submissiveness seems denied by their lack of passivity (item 26).

Groups C and D are indistinguishable with respect to their cluster scores: both have very low Ego Strength (scoring as low as group A), are considerably lower than either group A or group B on Social Adjustment, and show strikingly lower Emotional Expansiveness. However, despite this similarity, some differentiation between these two groups is suggested by residual items. While boys of both groups are far more passive in their approach to the environment (item 26) than any of the other boys, group C is more passive than group D and also more timid and fearful (item 31). Group D boys, on the other hand, show greater independence of thoughts and feelings (item 3), are more individualistic (item 15), seek more attention and praise (item 27), and are much more counterphobically fearless (item 29). Further, they are quite verbally expressive (item 25), much more so than any of the other groups. This, when combined with the finding that the mean mental test scores (Stanford-Binet, ages 6-15) of the three group D boys are among the four highest found in our total Guidance Study sample, suggests that they may qualify as the "brains" of their peer group. Groups C and D also exhibit different identification patterns: group C boys are more identified with their mothers (item 9); group D boys, with their fathers (item 10). Furthermore, only group C, among all groups, shows stronger identification with the mother than with the father. Taken together, these data suggest that, despite their similarity on cluster scores, the two groups may present different personality structures. Lack of emotional expansiveness, for example, may be for group C a direct expression of underlying passivity; for group D, it may reflect defensive reserve.

Girls. Figure 2 summarizes the median cluster scores for the four girls' ego control groups. (Again the median values for individual items in each cluster can be found in Table 1.)

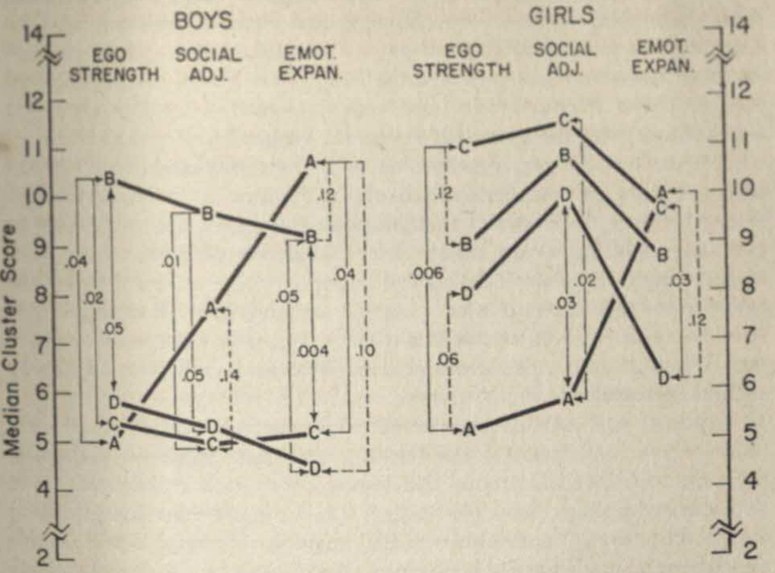

FIGURE 2—Median cluster scores for boys' and girls' ego control groups. Differences between groups significant at or below the .05 level are indicated by solid lines; probability levels between .05 and .15, by broken lines.

The girls' group A presents an undercontrolled pattern as pronounced as that of the boys'. Their low Ego Strength and high Emotional Expansiveness scores are strikingly similar to the boys' pattern, but, in sharp contrast, they score extremely low on Social Adjustment. An indication of the nature of their social behavior is provided by the residual items: while all other girls are uniformly low on aggressiveness (item 23) and nonconformity (item 15), group A girls characteristically score high on these two items. Furthermore, they are fearless (item 29), dominating (item 30), and not timid (item 31), although they lack independence of thoughts and feelings (item 3) and seek attention and praise (item 27). Consonant with the "acting-out" behavior shown by group A girls is their clear differentiation from other girls in showing less tendency to repress unacceptable impulses (item 12). Lastly, although these girls are strongly identified with their mothers (item 9), they show considerable conflict over their sexual identification (item 8).

Only group A girls show a distinctive cluster score profile; the remaining groups appear very similar, on residual items as well as on cluster scores. Despite this apparent homogeneity, it is possible to assign distinguishing characteristics to each of the groups, aided, in this instance, by an examination of *intra*cluster item differences among them.

Group C girls, who tend to be neither impulsive nor overcontrolled, show the highest degree of Ego Strength and Social Adjustment and are quite high on Emotional Expansiveness. On residual items, these girls are the most assertive (item 30) and the least timid (item 31) of the three groups. They differ further in being only moderately nurturant (item 28) —in contrast with the extremely nurturant group B and group D girls— while showing stronger identification with their mothers (item 9) than either of these groups. An analysis of the component items of the Ego Strength cluster demonstrates that, although they score uniformly high on each, their advantage over groups B and D derives primarily from three items: they appear more relaxed and easy going (item 11), they become ego involved in a variety of areas (item 5), and they are able to absorb their families' values without undue strain (item 13). The impression we gain, then, is that group C girls are characterized by considerable personal stability and interpersonal ease.

Group B girls are intermediate between groups C and D on all three of the cluster variables. Of the few characteristics which do distinguish them from the C and D groups, the clearest is their lack of independence of thoughts and feelings (item 3). Further, they are more repressive (item 12) and timid (item 31) and, although they share with groups C and D girls an extreme tendency toward conformity (item 15), group B girls experience difficulty in absorbing their families' cultural values without strain (item 13, Ego Strength cluster). Of all the girls' groups, they show the least identification with the father (item 10).

Only group D girls are clearly not undercontrolled; they also achieve the highest, although still only moderate, degree of overcontrol. They are best separated from all other groups by their very low Emotional Expansiveness. On both Ego Strength and Social Adjustment, group D girls are lower than those of groups B and C but still fall well within the middle range of the scales. The residual items permit little differentiation of these girls from the other groups. Group D girls, like those in groups B and C, are neither passive (item 26), individualistic (item 15), fearless (item 29), nor aggressive (item 23). They do seek attention and praise (item 27) and, in comparison with B and C girls, are somewhat less verbal (item 25). These characteristics, combined with the apparent stereotypy of their social behavior (items 18 and 22, Social Adjustment cluster) and their emotional constriction (item 32, Emotional Expansiveness cluster) suggest that group D girls adopt a "little lady" role which may serve a defensive function and, specifically, perhaps may relate to their conflict over sexual identification (item 8).

Discussion

While the personality sketches obtained through the ego control groupings are obviously incomplete and necessarily tentative, their feasibility supports the relevance of this parameter to personality organization in early adolescence. Furthermore, these syndromes encompass characteristics rather remote from those directly inferable from the ways in which the child manages his impulses during this period. For example, the rather "average" ego control pattern of group B girls cannot anticipate their distinct tendency to employ repressive defenses; the nature and the similarity of group C and group D boys' control scores cannot directly account for their opposite parent identification patterns.

Our use of two dimensions to characterize impulse control has been due as much to expediency as to any prior theoretical notions (*see* footnote 2). We find, however, that the two-variable control groupings have apparently allowed for greater discrimination among individuals than would have been suggested by a frequency distribution of a single measure of impulse control. From such a single score, for example, the distinction between boys' groups B and C, easily made from the scatterplot, would not have arisen and the rather suggestive differences between them would not have been detected. Certainly, it is not difficult—on a descriptive level, at least—to regard impulsivity as implying something more than the polar opposite of inflexibility. When we say of a person that he is "anything but" rigid and overcontrolled, we evoke an image of spontaneity but not necessarily one of extreme impulsivity. Nevertheless, the present results provide insufficient basis for a serious evaluation of the dimensionality of ego control as a theoretical construct; they do, however, indicate the promise of such an undertaking.

Any generalization of the results of this investigation to other developmental periods, or even to early adolescence experienced under other value systems, would seem to be a very risky endeavor. When the peer culture differs, or as it changes with approaching adult status, an adolescent's impulse control pattern may find different expressions and induce different social reactions. For example, boys' and girls' groups A share a pattern which links considerable impulsivity with emotional disturbance and extreme lability of expression. Yet, for boys, this syndrome permits an "average" level of social adjustment while, for girls, the resultant is pronounced social failure. This suggests that the prescriptions for socially appropriate behavior are widely divergent for boys and girls in early adolescent peer culture, specifically with respect to its sanctioning of impulsive behavior.

Unusually convenient support for this hypothesis is provided by Tuddenham (14) who conducted sociometric studies in the classrooms of these same children in the early school grades. He reports that in the fifth grade (approximately age 11) boys who are seen by their peers as "taking chances" also tend to be "popular" with their classmates and to be regarded as "leaders" of the group. Daring is not so rewarded for girls. Thus, the differing social adjustment of undercontrolling boys and girls may be tied to these contrasting peer evaluations, mediated by the differing self-perceptions they induce in the young adolescent. Furthermore, since these data also indicate that impulsivity in boys becomes steadily less valued through the school years, we may infer that group A boys will come under pressure to reappraise themselves and their behavior in relation to changing peer mores. Under such influences, the ego control pattern itself may be modified or, as the heterogeneity of peer culture increases with increasing age, the boy may tend to preserve his pattern by a self-selection of a more congenial social milieu which accepts, or even rewards, his impulsivity.

The two facets of impulse control appear to have somewhat different connotations for the two sexes. Examining the correlations between the two ego control items and the cluster scores over the total sample (Table 3), we find that boys' Ego Strength and Social Adjustment bear little relation to their degree of impulsivity but are substantially related to the inflexibility measure. For girls, the reverse conclusion is suggested: for them, degree of overcontrol is irrelevant while effective functioning tends to accompany a lack of impulsivity. A clue to an understanding of these relations is to be found in the interaction of the different social mores and maturational levels of boys and girls at this particular period. We may hypothesize that girls, who are biologically more mature at this age, are required to exert effective control over sexual impulses if they are to maintain satisfactory adjustment to self and to society. For boys, sexual undercontrol is less at issue while, on the other hand, excessive control of aggressive, assertive impulses is a social and personal handicap in early adolescence. From these premises the apparent greater importance of degree of impulsivity for girls, and degree of overcontrol for boys, to their respective personality organization may be deduced.

Apart from sex differences in rate of physical maturation during adolescence, variations within each sex on developmental status may be suspected to be of some relevance to ego control. A number of maturational indices were selected for comparing ego control groups and for correlation with the two control measures: age at maximum growth (MGA) and a "general factor" maturity score (10) for both sexes, age at menarche for girls, and age at reaching Sex Stage IV (7) for boys. For neither sex do the ego control groups differ significantly on any of the maturational indices. For boys, both control scores show near-zero correlations with each of the

TABLE 3

CORRELATIONS OF EGO CONTROL ITEMS WITH CLUSTER SCORES

	EGO STRENGTH		SOCIAL ADJUSTMENT		EMOTIONAL EXPANSIVENESS	
Item	*Boys*	*Girls*	*Boys*	*Girls*	*Boys*	*Girls*
Is undercontrolling, acts upon impulse	—.13	—.24	.26	—.24	.46	.27
Is rigid, overcontrolling, inflexible	—.48	.02	—.57	.06	—.82	—.27

three maturational indices, but, for girls, there is the curious finding that MGA and menarcheal age, despite their strong positive intercorrelation ($r = .71$ for all Guidance Study girls), show substantial and *opposite* correlations with degree of overcontrol (MGA, *gamma* = .40; menarche, *gamma* = −.31). (The same reversal, but with lower values, is found for undercontrol.) Inspection of scatterplots reveals that girls whose growth spurt is relatively early (before age 11) and those whose menarche occurs later than average (after age 13) tend towards extreme overcontrol of impulses. The remainder of the range on both maturational indices shows no relation to either control item. This suggests that the interval between MGA and menarche may relate to the rigidity and inflexibility of impulse control. We find that girls for whom MGA precedes menarche by two or more years have a mean overcontrol score of 12.6 (7.8 on the impulsivity measure); for those who reach both of these developmental points within the same year, the corresponding values are 7.8 and 12.5. For the remaining girls, with the modal interval of one to two years, the mean scores are intermediate: 10.1 and 10.3. While these three groups vary somewhat, and in the obvious directions, on mean age for MGA and menarche, it seems more probable that it is the timing relation between rate of growth in size and sexual maturation which underlies their variation along the ego control dimensions. Why a relatively tall and sexually immature girl should tend toward greater inflexibility in handling her impulses during early adolescence than one who is small and relatively mature—if, indeed, this proves to be a replicable finding—is difficult to understand. What it does argue, however, is that, despite the existence of a strong general factor coordinating the various processes of maturation during adolescence, there may be useful information in the temporal patterning of these processes. If this is so, then —to take one example—investigations of personality correlates of maturational rate may be made more powerful by a consideration of the possible impact of asynchronies among the various components of development in adolescence.

The present analysis of the behavioral and intrapsychic correlates of ego control is thus qualified by its limitation to a given period of development (and the different maturity levels of the two sexes during that period) and to the special qualities of early adolescent culture. Specifying the link between what we have found for this period and the picture which will be discernible in adulthood—whether it will indicate constancy, predictable transformations, or puzzling discontinuities—must await the completion of a current assessment of these persons, now 30 years of age.

SUMMARY

Samples of children in their early adolescence (20 of each sex) were partitioned into eight ego control groups on the basis of clusters found visually in bivariate distributions of ratings of overcontrol and undercontrol of

impulses. Differences among these groups on three primary variables (Ego Strength, Social Adjustment, and Emotional Expansiveness) were found which, when combined with group differences occurring in a supplementary set of personality evaluations, permitted the assignment of partial and tentative personality syndromes to each of the ego control patterns. Other parameters, such as the sex of the child and differential maturational rates, determine the specific natures of the relations between control mode and personality organization.

REFERENCES

1. BLOCK, J. The development of an MMPI-based scale to measure ego control. Institute of Personality Assessment and Research Report, 1953 (mimeo).
2. BLOCK, J. *The Q-sort method in personality assessment and psychiatric research.* C. C. Thomas, 1961.
3. BLOCK, J., & THOMAS, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.,* 1955, 51, 254-259.
4. BLOCK, JEANNE, & MARTIN, B. Prediction of behavior of children under frustration. *J. abnorm. soc. Psychol.,* 1955, 51, 281-285.
5. FREUD, S. *An outline of psychoanalysis.* Norton, 1949.
6. GOODMAN, L. A., & KRUSKAL, W. H. Measures of association for cross-classifications. *J. Amer. statist. Ass.,* 1954, 49, 732-769.
7. GREULICH, W. W., *et al.* Somatic and endocrine studies of puberal and adolescent boys. *Monogr. Soc. Res. Child Develpm.,* 1942, 7, No. 3 (Serial No. 33).
8. KALIS, B. L. Testing hypotheses with narrative case material. Paper read at 14th V. A. Regional Psychology Research Conference of the San Francisco Bay Area, Berkeley, July, 1956.
9. MACFARLANE, J. W. Studies in child guidance: I. Methodology of data collection and organization. *Monogr. Soc. Res. Child Develpm.,* 1938, 3, No. 6 (Serial No. 19).
10. NICOLSON, A. B., & HANLEY, C. Indices of physiological maturity: derivation and interrelationships. *Child Develpm.,* 1953, 24, 3-38.
11. REDL, F., & WINEMAN, D. *Children who hate.* Free Press, 1951.
12. ROSENTHAL, I. Sampling distribution of the nonparametric correlation coefficient, gamma. Paper read at Western Psychol. Ass., San Diego, April, 1959.
13. SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, P. S. Some childrearing antecedents of aggression and dependency in growing children. *Genet. Psychol. Monogr.,* 1953, 47, 135-234.
14. TUDDENHAM, R. D. Studies in reputation: III. Correlates of popularity among elementary school children. *J. educ. Psychol.,* 1951, 42, 257-276.

LEVEL OF STIMULUS PRETRAINING AND PAIRED-ASSOCIATE LEARNING[1]

HAYNE W. REESE*
University of Buffalo

Previous studies of the effects of varying amounts of stimulus pretraining on subsequent discrimination or paired-associate learning have yielded inconclusive results. Some reported significantly more facilitation following greater amounts of pretraining, but did not control for the effects of possible performance set differences in groups given different amounts of pretraining (e.g., 2, 5). Others controlled performance set, but found that different amounts of pretraining did not produce significantly different amounts of performance set or of acquired distinctiveness (1, 3). The *S*s, stimuli, and task I responses were comparable in these studies, but the former studies used button-pushing or switch-throwing responses in task II, whereas the latter studies used the Star Discrimeter. In the present study, stimuli and responses comparable to those used in the former studies were used, but children were used as *S*s because it was desired to simplify the tasks and to minimize the probability that *S*s would have preexperimentally learned names for the stimuli used in the condition which controlled performance set effects.

Two specific predictions were tested: (a) the higher the level of pretraining, the better the performance on task II will be; and (b) the higher the level of pretraining, the greater the difference between the experimental (acquired distinctiveness) and performance-set control conditions will be.

In task I, *S*s learned to associate distinctive nonsense syllables with two stimuli. Three groups were trained to a low, moderate, or high criterion.

---

* Department of Psychology, University of Buffalo, Buffalo 14, New York.

[1] This research was conducted while the writer was a research assistant at, and with funds made available by, the Iowa Child Welfare Research Station. The writer is indebted to Dr. Charles C. Spiker for his aid in the research.

Child Develpm., 1961, 32, 89-93.

In task II, *Ss* learned to associate button-pushing responses with these two stimuli and two others. The two stimuli used in both tasks constitute the experimental condition, and the two used only in task II constitute the control condition. Presumably, performance set affects performance on the control condition, and both performance set and acquired distinctiveness affect performance on the experimental condition. Therefore, the difference between performance on the two conditions can be attributed to acquired distinctiveness, with performance set effects controlled.

METHOD

Subjects

The *Ss* were 90 fourth and fifth grade school children from two public schools in Ottumwa, Iowa.[2] They were assigned to the experimental groups in a random manner.

Apparatus

The apparatus was essentially the same as that described in detail by Spiker and Holton (6), except that the response unit contained only two push-buttons, located side-by-side approximately 5 in. apart. There was a small pilot lamp above each button.

The stimulus unit contained a single stimulus aperture through which lights differing in hue could be presented one at a time. Corning Glass Co. filters were used to produce four stimuli: two red lights, a light red (No. 2418) and a dark red (No. 2030); and two green lights, a light green (No. 3718 and No. 4784 together) and a dark green (Nos. 3384, 3389, and 5031 together).

The task I response words were three-letter nonsense syllables, *lev* and *wog*. In terms of common elements and speech sounds the syllables are highly distinctive. The syllables were printed on 3 in. by 5 in. cards, which were presented manually by *E*.

Procedure

Pretraining task. The response unit was concealed throughout task I. The *Ss* learned to associate the nonsense syllables with two stimuli in task I. Group H was trained to a high criterion, 19 correct responses in 20 consecutive stimulus presentations; group M was trained to a moderate criterion, seven correct responses in 10 consecutive presentations; and group L was trained to a low criterion, three consecutive correct responses. The red stimuli were used for half of each group, and the green stimuli for the other half. There were 30 *Ss* in each of the three experimental groups.

---

[2] The writer is indebted to Mr. R. O. Wright, Director of Curriculum, Ottumwa Public School System, Ottumwa, Iowa, for his generous permission to use the *Ss* of this experiment. Thanks are also due Mr. Glenn J. James, Principal, Hedrick and Fairview Elementary Schools, and his teaching staff, for their assistance in many ways.

The anticipation time was 2 sec., and the end of this interval was signalled by a bell. When *S* gave an incorrect response, or failed to respond, he was shown the card bearing the correct syllable during a 3-sec. joint-presentation interval and was required to read it aloud. The interstimulus interval was 2 sec.

Transfer task. Immediately after *S* reached the appropriate task I learning criterion, the response unit was placed in position in front of the stimulus-exposure unit and below the stimulus aperture. All four stimuli were used in task II; *Ss* learned to associate light green and dark red with the left button, and dark green and light red with the right button. The pilot lamp above the correct button came on at the end of a 3-sec. anticipation time and remained on during a 2-sec. joint-presentation interval. The interstimulus interval was 2 sec. The *Ss* were instructed to respond prior to the onset of the pilot lamp and were allowed only one response to each stimulus presentation. The bell rang immediately after correct responses.

All *Ss* were given 18 trials, each trial consisting of the presentation of all four stimuli once each. No stimulus was repeated consecutively, and no stimulus alternated with one of similar hue more than once.

RESULTS

The mean numbers of stimulus presentations through criterion in task I were: group L, 6.1 (SD = 3.2); group M, 9.3 (SD = 3.3); and group H, 32.3 (SD = 21.5).

The mean numbers of errors in blocks of six trials on task II are given in Table 1. Analysis of variance of these data indicated that the difference between the over-all means for the experimental and control conditions was significant ($F = 4.61$; $df = 1, 87$; $p < .05$). This effect did not interact

TABLE 1

MEAN NUMBER OF ERRORS ON TASK II

Condition	Blocks of Six Trials 1	2	3	All Trials
Group L				
Control	5.6	4.1	4.3	4.7
Experimental	5.3	3.9	3.1	4.1
Group M				
Control	5.4	4.2	3.2	4.3
Experimental	4.6	3.5	3.4	3.8
Group H				
Control	5.5	2.7	3.0	3.7
Experimental	3.8	3.1	2.4	3.1

significantly with Levels of Pretraining ($F < 1.00$), but the triple interaction of Conditions, Levels of Pretraining, and Trials was significant ($F = 2.84$; $df = 4, 174$; $p < .025$). The simple effects involved in this interaction were tested by analysis of variance of the data for each block of trials separately. These analyses revealed only one significant effect; in the first six trials the Conditions effect was significant ($F = 9.86$; $df = 1, 87$; $p < .005$), indicating that performance on the experimental condition was significantly superior to that on the control condition only in the first block of trials. Examination of the means on the first block of trials shows that the difference between the control and experimental conditions is much greater in group H than in group L, but, since Conditions did not interact significantly with Level of Pretraining ($F = 1.74$; $df = 2, 87$; $p > .10$), this effect was not statistically reliable.

Discussion

The analyses indicated that the facilitation attributable to acquired distinctiveness was significantly greater than that attributable to performance set only in the first block of trials. This finding would be expected on the assumption that the experimental and control learning curves approach the same asymptote, so that differences become smaller as learning progresses.

Contrary to the second prediction, there was no significant increase in acquired distinctiveness with increased level of pretraining, suggesting that acquired distinctiveness is maximal after very low levels of pretraining, about six stimulus presentations in the present study. None of the theoretical mechanisms which have been suggested account for the rapidity of the development of the effect. Those based on learned mechanisms would seem to require a greater number of pretraining trials to permit learning to occur, whereas those based on perceptual responses do not lead to the prediction of a difference between the experimental and control conditions when these conditions are within-*S*s factors and the stimuli within each condition differ on the same dimensions, e.g., saturation and brightness in the present study.

The data suggested that performance set may have increased with increasing level of pretraining, and, although the analyses of variance did not support this suggestion, *t* tests of the differences between groups L and H (the extreme groups on the independent variable) offered some support, in that the difference in the second block of trials approached significance ($t = 1.90$; $df = 87$; $p = .06$), and the difference over all trials was significant ($t = 2.06$; $df = 87$; $p < .05$). These *t* tests do not provide statistical justification for an assertion that performance set must be controlled in studies of acquired distinctiveness, especially since previous studies which controlled performance set found that it did not increase significantly (1, 3, 4), but the data of these and the present study suggest that such a trend may be present and that performance set should be controlled.

SUMMARY

It was predicted that (a) the higher the level of stimulus pretraining, the better the performance on a motor paired-associate transfer task will be; and (b) the higher the level of stimulus pretraining, the greater the difference between an acquired distinctiveness condition and a performance set control condition will be.

Ninety fourth and fifth grade school children learned distinctive nonsense-syllable names for two similar stimuli in task I. Group L learned the associations to a low criterion; group M, to an intermediate criterion; and group H, to a high criterion. In task II, *Ss* learned to associate button-pushing responses with these stimuli and two others.

The first prediction was supported only by a test of the difference between the task II performance of groups L and H, which suggested the presence of a trend. The second prediction was not verified; the difference between the experimental and control conditions was significant in the first block of trials, indicating that the facilitation on the experimental condition was attributable not only to performance set but also to acquired distinctiveness, but this effect was not significantly different in the three groups.

REFERENCES

1. CANTOR, J. H. Amount of pretraining as a factor in stimulus predifferentiation and performance set. *J. exp. Psychol.,* 1955, 50, 180-184.
2. GAGNE, R. M., & BAKER, K. E. Stimulus predifferentiation as a factor in transfer of training. *J. exp. Psychol.,* 1950, 40, 439-451.
3. MCCORMACK, P. D. Negative transfer in motor performance following a critical amount of verbal pretraining. *Percept. mot. Skills,* 1958, 8, 27-31.
4. REESE, H. W. Motor paired-associate learning and stimulus pretraining. *Child Develpm.,* 1960, 31, 505-513.
5. ROSSMAN, I. L., & GOSS, A. E. The acquired distinctiveness of cues: the role of discriminative verbal responses in facilitating the acquisition of discriminative motor responses. *J. exp. Psychol.,* 1951, 42, 173-182.
6. SPIKER, C. C., & HOLTON, R. B. Associative transfer in motor paired-associate learning as a function of amount of first-task practice. *J. exp. Psychol.,* 1958, 56, 123-132.

MEASUREMENT OF MOVEMENT IN THE HUMAN NEWBORN: A NEW TECHNIQUE[1,2]

WILLIAM KESSEN,* LOUISE S. HENDRY, *and*
ANNE-MARIE LEUTZENDORFF
Yale University

The behavior of the human newborn shows little of the intentional character of later infant or adult behavior; that is, the responses of the newborn can only tenuously and with difficulty be related to the events in his environment. The newborn behaves without goals (10). Largely for this reason, no generally accepted taxonomy of responses has been developed for neonatal behavior, and researchers have used a wide range of techniques in their attempts to make sense of the disconcerting variety of neonatal movement. The most highly regarded techniques used in the last several decades have aimed at description of "mass activity" or "general activity" rather than at the description of circumscribed local movements. Even with this limitation on the specification of the response-class to be studied, however, there has been a great deal of diversity in technical methods for recording activity or movement. The human observer has been used, without instrumentation other than pencil and paper, to catch the complexities of neonatal behavior (16), although the more common employment of the gadget-free human observer has been restricted by checklists or rating scales (3, 4, 14). Other than the unassisted human being, the technique of choice in the recording of neonatal movement has been one form or another of the stabilimeter (3, 6, 11, 14). This ingenious device, closely akin to mechanisms used to assess activity in animals (e.g., 2), provides a numerical index of

---

* Department of Psychology and Child Study Center, Yale University, 333 Cedar Street, New Haven, Connecticut.

[1] The research reported here was supported in part by Research Grant M-1787, United States Public Health Service.

[2] The writers express their thanks to Dr. Milton J. E. Senn and to the staff of Grace-New Haven Community Hospital for their generous cooperation in support of the research. E. J. Williams and J. P. Williams assisted in the collection of data and their analysis.

Child Develpm., 1961, 32, 95-105.

the infant's movement, an index which has been shown to be sensitive to environmental variations of several kinds (Irwin, *passim*). The advantages of stabilimetric techniques over the unaided human observer in a quantitative treatment of newborn movement are clear: The danger of observer bias and contamination is reduced markedly; there is no shifting focus on different limbs; the baby's movements can be recorded over long stretches of time; and there is relatively little ambiguity in scoring the data-protocols. This is not to say that the human observer is an obsolete device in the study of neonatal behavior. At current levels of ignorance about what makes the neonatal organism work, it is of crucial importance that the sensitive human instrument be used to develop hunches, to make preliminary confirmations of these hunches in general unstructured observation, and, in the case of well-developed checklists or rating scales, to test hypotheses systematically. Within the limits of these qualifications, objective and reproducible techniques for recording movement will have substantial value for our understanding of the psychology of the neonate.

There has been available for many years a recording technique which would seem at first glance to be ideally suited to a study of neonatal movement—the motion-picture camera. Surprisingly enough, although films of young children have often been made (7, 9, 13), these films have only rarely been put to quantitative analysis (17). Gesell and Halverson (8) made systematic cinematographic observation of a single child over a period of several months, scoring for posture and laterality. Gilmer (9) has used film to describe in greater detail certain neonatal movements, and Birdwhistell (e.g., 1) has presented a general system of movement analysis which may have relevance to the study of infants. These studies appear to be fair representatives of the research use of cinematographic techniques in the study of young infants. It may be that motion-picture photography is such an easy observational technique compared with alternative methods that the researcher collects thousands of feet of film with understandable gusto, only much later to face the question: What do I do with all these pictures? Descriptive and illustrative uses of film have been the most productive answers to date.

As part of an investigation of neonatal behavior, we set out to develop a technique which would permit the assignment of a numerical index to the movement of newborns. Furthermore, we wanted a technique which would have the advantage over stabilimetric devices on two counts: (a) in providing a fully comparable score, that is, a movement index that could be used in different research settings without introducing problems of mechanical comparability of instruments; and (b) in preserving a record of moment-to-moment changes in the behavior of the baby. For these purposes, motion-picture photography seemed appropriate; our problem was to devise a technique of film-analysis which would be reproducible, fully comparable across different settings, and productive of a meaningful measure.

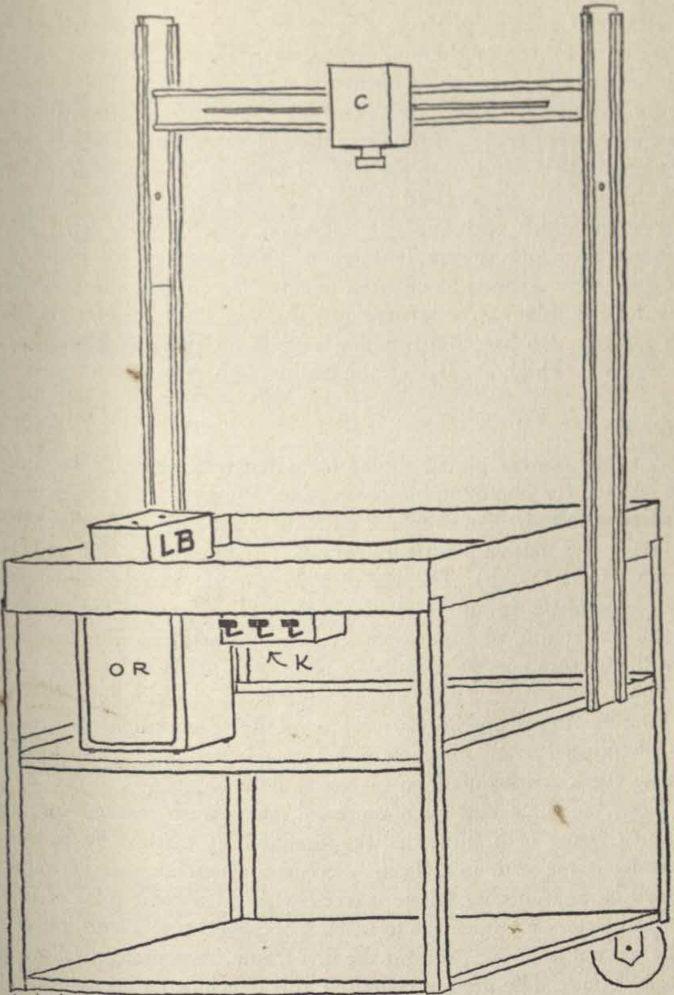

FIGURE 1—Observation cart in which infants were photographed. C: camera; K: telegraph keys; OR: Esterline-Angus operations recorder; LB: light-box used to synchronize OR and film.

METHOD

Setting and Subjects

Normal, full-term infants delivered in the maternity service of Grace-New Haven Community Hospital were observed during the lying-in period of five days. Fifteen of these newborns were photographed on each day of their hospital stay, 10 of them for five 30-sec. film strips daily, the remaining five for two 30-sec. strips daily. The babies were brought from the nursery to an unoccupied room on the same floor of the hospital, where they were photographed shortly before the midday feeding. Throughout the observation, the infants were dressed in hospital shirt and diapers. The period of observation, which included data-collection described in detail elsewhere (10), was from 20 to 30 min. in duration. The five 30-sec. filmstrips made for 10 babies were separated one from another by 90 sec.; the two filmstrips made for five babies were separated by 450 sec. Each strip was marked for later identification by photographing a series of letters and numbers coded to identify the child, the day of observation, and the strip.

Apparatus

The infants were placed supine on a mattress which fit an equipment cart adapted for photographic observation. Figure 1 is a schematic representation of this cart. An overhead girder approximately 3 ft. above the baby supported the motion-picture camera C (in this case, Bell and Howell 16-mm, Model DL-70). The use of a wide-angle lens (10-mm) provided a full view of the infant at relatively close placement of the camera. The camera was run at 16 frames per sec., and no artificial illumination other than the overhead lights already in place was used. A bank of telegraph keys (K) on each side of the cart permitted the recording of other data on an Esterline-Angus operations recorder (OR). Depression of certain keys also illuminated small indicator lights in a box (LB), which lights were used to check synchronization of film and recorder.

When the films had been processed, they were marked for analysis. The first frame of a filmstrip was located and marked by painting the preceding frame with ink. Then, a predetermined number of frames was counted off and another frame marked with a different color of ink. The most satisfactory method was to mark successive frames with ink of colors in the spectral order, i.e., red for the first frame, then orange, yellow, green, blue, and violet. The present technique was developed with a 5-sec. interval between frames selected for analysis (80 frames at 16/sec.), but other intervals can be used [*see* (15) for an analysis using a 1-sec. interval]. Thus, after the marking was completed, six frames from each 30-sec. filmstrip were marked for analysis—the first frame, the 81st (5 sec. from the start of the strip), the 161st (10 sec.), the 241st (15 sec.), the 321st (20 sec.) and the 401st (25 sec. from the beginning of the filmstrip).

Technique of Film-Analysis

The next step toward obtaining a reproducible measure of neonatal movement was to make a permanent copy of the filmframes selected for analysis. This step was taken by means of the filmanalyzer illustrated schematically in Figure 2. The original filmstrip (usually on a 100-ft. roll of film with several other strips) was placed on a reel (RI) and fed past a projector-lamp box (P) to another reel. This second reel was driven by a sewing machine motor (M) through a variable speed control device (R), so that a foot switch operated by the observer controlled the rate at which the film moved past the projector-lamp box. When a film frame selected for analysis came into position opposite the lamp, the film drive was stopped. The image of the infant was projected against a first-surface mirror (Mi) in a wooden chimney (C) and thence to a glass surface (G), 11 by 14 in.

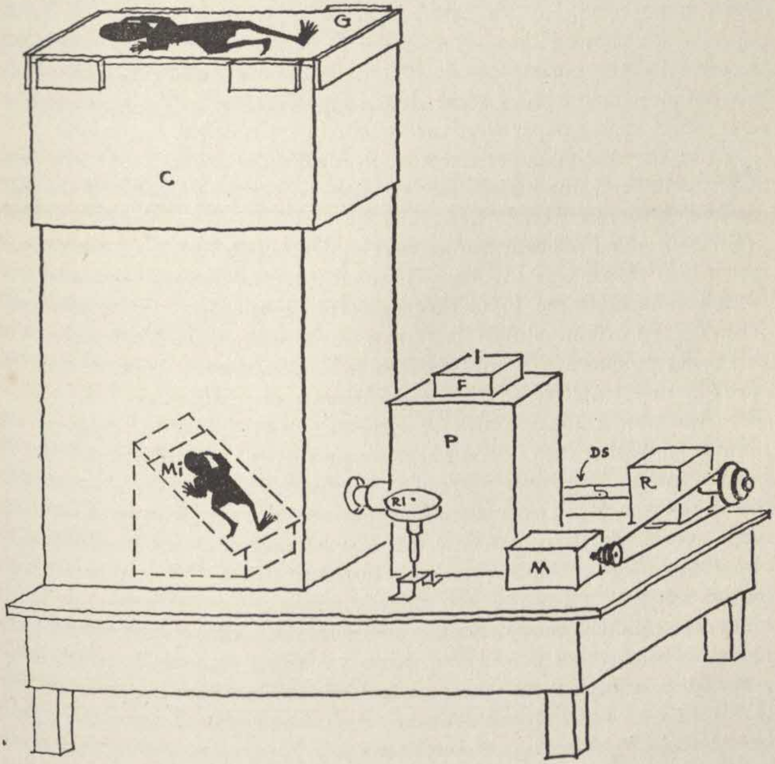

FIGURE 2—Schematic representation of filmanalyzer used in the present study. M: motor; R: variable-speed device; DS: driveshaft to takeup reel; RI: film reel; P: projector-lamp box; F: fan; Mi: first-surface mirror; C: enclosed wooden chimney; G: plate glass sheet.

in size. Draftsman's tracing paper was placed on this glass surface, and significant details of the image were traced for analysis. It should be noted once more that the particular technique for analysis and the particular dimensions of the apparatus used are not crucial in determining the measure of movement under study. Calibration procedures described below permit precise comparison of film analyses performed on instruments quite different from the one illustrated here.

Each of the filmframes selected for analysis was projected in the manner just described, and the position of the following parts of the infant's body was indicated by dots marked on tracing paper: visible ear, right palm, left palm, right knee, left knee, right ankle, left ankle. It was necessary also to mark on each sheet of tracing paper certain "landmarks," (usually the light box), so that one sheet could be accurately superimposed on another. Thus, for each filmframe selected for analysis (six per 30-sec. observation period, in this case), there was prepared a sheet of tracing paper which showed dots representing the position of seven body parts. Occasionally, the outline of the infant's body was sketched; occasionally, this outline was made in some detail. However, only the seven dots on each seet of tracing paper were used to obtain the movement measure.

Once the tracing-paper records of filmframes were made, the next task was to get a numerical index of changes in position of the body parts marked. A summary sheet was drawn, in which the dots representing limb positions from all relevant frames were shown, again coded in color to indicate which frame was involved. Then, the linear distance was measured from ear-mark (frame 1) to ear-mark (frame 81), from right-palm-mark (frame 81) to right-palm-mark (frame 161), and so on. Measurements of this order produced the protocol from which movement measure was extracted. An example of the protocol is shown in Table 1. As can be seen, there was now available an *index of displacement* for several parts of the infant's body, that is, a numerical representation of the shift in position of head, arms, and legs over a specified interval. It is important to note that this is not a measure of total movement, but rather a measure of displacement over a predetermined interval. The neonate may show wild flailing of arms and legs in the interval, only to return to exactly the same position by the time the next frame selected for analysis rolls by. We felt this possibility was unlikely enough to proceed with the development of the technique. Of course, shorter and shorter intervals could be used (e.g., examining every frame at high-speed photography) with an increase in approximation to a measure of total movement; correspondingly, the analytic task could be shortened by choosing a longer interval (e.g., a frame-sketch taken every 15 sec.). Our exploratory studies had led to a choice of the 5-sec. interval as optimally balanced between hopelessly difficult isomorphic representation of total movement and an impressionistic determination.

Thus, at the level of data-protocols, our index of displacement is reduced from "the real child" in three important ways: (a) it is based on film from

TABLE 1

REPRESENTATIVE EXAMPLE OF BASIC PROTOCOL

Subject A-1, *Day* 1, *Observation* 1 *Analyzer:* ALS

	FRAME-TO-FRAME DISPLACEMENT OF IMAGE (cm.)				
	1–81	81–161	161–241	241–321	321–401
RH	.75	.65	.85	1.65	.90
LH	.30	6.15	.95	5.70	1.05
RK	3.05	1.70	.70	2.00	1.40
LK	3.40	1.25	4.30	3.10	1.95
RA	2.30	.60	.75	2.20	.60
LA	2.20	1.15	3.15	1.35	.40
(L)E	.10	.80	.30	.10	.75

NOTE.—R indicates *right*; L, *left*; H, *hand*; K, *knee*; A, *ankle*; E, *ear*.

a single fixed-position camera; (b) it involves only seven points on the body; (c) it is an index of displacement of parts, not of total movement.

Even the reduced protocol index is not easily manageable for all purposes, however, and for ease of computation, two further reductions were made. Displacements over the entire 30-sec. observation period were averaged to obtain a mean index of displacement, and a single index of movement per observation was obtained by adding the two palm and the two ankle measuses. Therefore, in most of the findings reported below, a single number per child per observation period is used, and it is this numeral which will be called the Movement Index (MI). More detailed examinations of displacement can be made, however, from the basic protocol illustrated in Table 1.

Calibration

In order to ascertain the true distance of displacement, a calibration was made of the camera-rig and filmanalyzer. A black paper strip, 50 cm. in length, was photographed in the position normally occupied by an infant on the observation cart, and the processed film of this strip was projected through the film analyzer and measured. The obtained length of the projected image of the strip was 18.30 cm., so that the protocol data need only to be multiplied by 50/18.30 or 2.73 to obtain a measure of true horizontal displacement of limbs over a 5-sec. interval. The resulting metric, an index of displacement in centimeters per limb over a 5-sec. period (cm./limb/5-sec.), although of no particular interest in itself, has the inestimable value of permitting numerical comparison of results from one research setting to another.

Reliability

The entire analytic procedure, from the marking of film to the presentation of a summary protocol, was performed on a randomly selected strip by two observers. The interobserver coefficient of agreement (*r*) was .96.

APPLICATION OF THE METHOD

The method of filmanalysis described in the foregoing pages has been and is being used in several studies of neonatal behavior, where its usefulness is being put to the central test, i.e., whether meaningful statements can be developed using it. Some of the results obtained will be reviewed here, together with some previously unreported findings, in order to demonstrate the flexibility and appropriateness of the MI measure.

Individual Differences

Stable individual differences over the five days of the lying-in period have been demonstrated for the movement measure, even when only one 30-sec. filmstrip per day was used in the analysis (10).

Day-to-Day Variation

The MI measure showed systematic and significant increases over the five-day lying-in period, confirming earlier observations of an increase in activity over the first days of life (10).

Laterality Variation

One of the advantages of a cinematographic technique is that it permits the accurate assessment of variations in laterality, a problem that has been of interest to psychologists of infancy for some time (e.g., 12). Figure 3 shows the values of MI for right and left palm, for right and left ankle, and for these values summed, over the lying-in period. As can be seen, there is no systematic superiority of one side over another during these first days of life. Nor does this representation of 15 infants blur underlying stable individual differences in sidedness. For each of the five days of observation, two inequalities can be determined—right palm displaced more or less than left palm, and right ankle displaced more or less than left ankle. Thus, for the entire period of study, each child will show a "sidedness" score from 0 (no instances of right palm or ankle displaced more than left palm or ankle) to 10 (right member displaced more than left in every instance). The obtained array of sidedness scores can be evaluated against a hypothesis of random variation of sidedness by matching the obtained results against expected values drawn from a binomial expansion. Table 2 shows the result of this comparison. It seems reasonable to conclude from this observation that newborn infants show no systematic variation in laterality of movement during the first five days of life.

The null hypothesis with respect to laterality will not hold if head position is examined. The MI tracings were scored for position of the head,

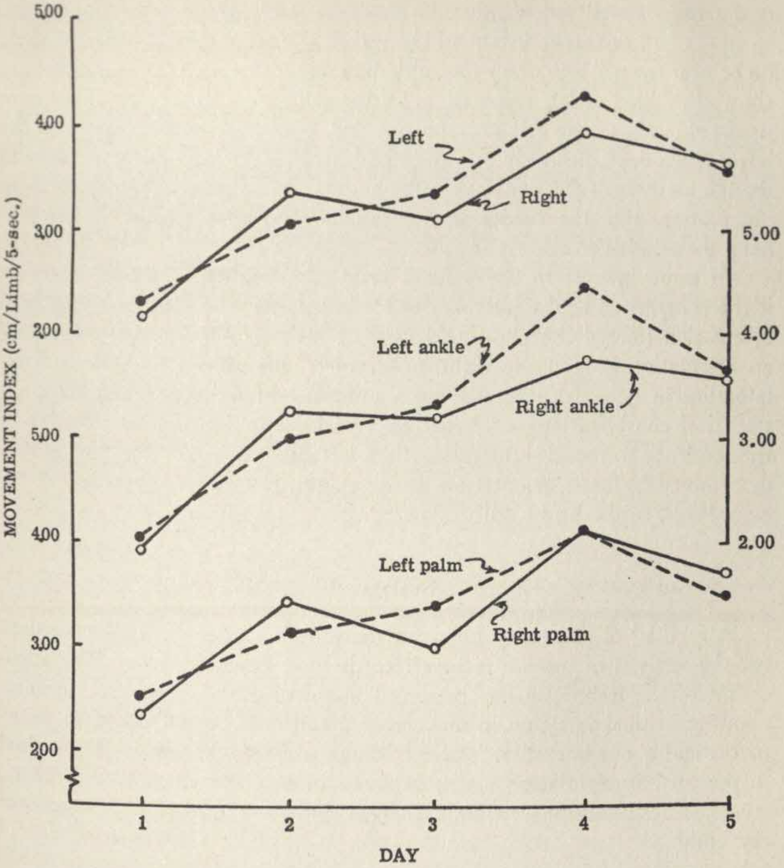

FIGURE 3—Displacement of limbs over the first five days of life. Lower curves show movement of palms; middle curves show movement of ankles; upper curves show the summation of the lower pairs of curves.

TABLE 2

NUMBER OF INFANTS SHOWING THREE DEGREES OF LATERALITY (OBTAINED) AND NUMBER TO BE EXPECTED ON A HYPOTHESIS OF RANDOM DISTRIBUTION OF PREFERENCE FOR RIGHT OR LEFT

	SIDEDNESS INDEX		
	0 1 2 3	4 5 6	7 8 9 10
Expected	2.6	9.8	2.6
Obtained	4	9	2

and a head-laterality determination made for each day of the hospital stay. In nine of 15 cases, the infant's head was to the right side (i.e., right cheek on or near the pad) *on every day,* although one of these children occasionally showed midline or left orientation. Of the remaining six infants, four were predominantly to the left only on one day, one was predominantly to the left on two days, and only one child had his head turned predominantly to the left on three of the five days in the hospital. We may conclude from these observations that the babies in our sample showed a strong tendency to keep their heads to the right.

Of more interest to the general issue of laterality variation, however, is the comparison of the laterality of MI in infants who show a strong head orientation toward the right with that of infants who show not-so-strong an orientation toward the right. Contrary to any simple hypothesis about laterality in neonates, there is no significant difference (either based on statistical considerations or based on visual inspection of the records) in the tendency to move right rather than left limbs between the two groups determined by head orientation. It is apparently not the case that, as the head leads, so the limbs will follow.

SUMMARY

A method of observing newborn movement and a technical procedure for the analysis of motion-picture records have been described. The application of the technique has produced interesting and suggestive findings about individual variation in movement, about day-to-day changes in movement, and about laterality. These findings support the use of the method in the further exploration of the behavior in newborns, particularly in the study of response to systematic changes in the external environment of the child.

REFERENCES

1. BIRDWHISTELL, R. L. Background to kinesics. *Etc. Rev. gen. Semant.,* 1955, 13, 10-18.
2. CAMPBELL, B. A. Design and reliability of a new activity-recording device. *J. comp. physiol. Psychol.,* 1954, 47, 90-92.
3. ESCALONA, S. K., & LEITCH, M. Early phases of personality development: a non-normative study of infant behavior. *Monogr. Soc. Res. Child Develpm.,* 1952, 17, No. 1 (Serial No. 54).
4. FRIES, M. E. Psychosomatic relationships between mother and infant. *Psychosom. Med.,* 1944, 6, 159-162.
5. IRWIN, O. C. The amount and nature of activities of newborn infants under constant external stimulating conditions during the first ten days of life. *Genet. Psychol. Monogr.,* 1930, 8, 1-92.

6. Irwin, O. C. The distribution of the amount of motility in young infants between two nursing periods. *J. comp. Psychol.,* 1932, 14, 429-445.

7. Gesell, A. *An atlas of infant behavior.* Yale Univer. Press, 1934.

8. Gesell, A., & Halverson, H. M. The daily maturation of infant behavior. A cinema study of postures, movements and laterality. *J. genet. Psychol.,* 1942, 61, 3-32.

9. Gilmer, B. v.H. An analysis of the spontaneous responses of the newborn infant. *J. genet. Psychol.,* 1933, 42, 392-405.

10. Kessen, W., Williams, E. J., & Williams, J. P. Selection and test of response measures in the study of the human newborn. *Child Develpm.,* 1961, 32, 7-24.

11. Marquis, D. P. Can conditioned responses be established in the newborn infant? *J. genet. Psychol.,* 1931, 39, 479-492.

12. Marquis, D. P. A study of activity and postures in infants' sleep. *J. genet. Psychol.,* 1933, 42, 51-69.

13. Murphy, L. B. *Personality in young children.* Vol. 1. Basic Books, 1956.

14. Pratt, K. C., Nelson, A. K., & Sun, K. H. The behavior of the newborn infant. *Ohio State Univer. Stud., Cont. Psychol.,* 1930, No. 10.

15. Williams, J. P., & Kessen, W. Effect of hand-mouth contacting on neonatal movement. *Child Develpm.,* in press.

16. Wolf, K. M. The origins of individuality. Mimeographed book, 1957.

17. Wright, H. F. Observational child study. In P. H. Mussen, (Ed.), *Handbook of research methods in child development.* Wiley, 1960. Pp. 71-139.

AMPLITUDE OF AN INSTRUMENTAL RESPONSE FOLLOWING THE CESSATION OF REWARD[1]

RUTH B. HOLTON*
State University of Iowa

Several investigators (1, 4, 6, 7, 8) have reported increments in some measures of response strength following the omission of a customary reward. Few studies, however, have demonstrated variations in the magnitude of these increments as a function of experimental variables.

Roussel (7) reports data which suggest that amount of previous training is positively related to increased running speed in rats following nonreward. Haner and Brown (3) found children exerted more force when blocking was introduced near a goal than when it occurred at points farther removed from the goal. These findings suggest that increments in the amplitude of an instrumental response, occurring as a consequence of nonreinforcement, may be affected by level of previous reinforced training and distance of the point of nonreward from the goal. The study reported here was designed to test these hypotheses with young children.

A marble reinforcement was delivered each time *S* responded to (pushed) the correct one of two stimulus windows. The number of times *S* pushed the

---

* Deceased.

[1] This article is based upon a Ph.D. dissertation submitted to the faculty of the Iowa Child Welfare Research Station, State University of Iowa. The author is indebted to Dr. Charles C. Spiker for directing the research and to Mr. John Peterson for constructing the apparatus.

Child Develpm., 1961, 32, 107-116.

correct stimulus window constituted level of reinforced training. Distance from the goal was expressed in terms of the number of marbles necessary for *S* to finish filling a marble board, the completion of which entitled him to a toy reward. Marble reinforcements were withheld after two levels of reinforced training and at two distances from the goal.

It was expected that, with goal distance held constant, *S*s having a relatively high number of reinforced trials prior to the failure of reward would show greater increments in response amplitude than *S*s having fewer reinforcements. Furthermore, with level of reinforcement constant, it was expected that *S*s nonrewarded close to the goal would show greater increments in response amplitude than *S*s nonrewarded far from the goal.

METHOD

Subjects

Forty-five children, ranging in age from 3 years, 8 months, to 5 years, 6 months, drawn from the Iowa Child Welfare Research Station Laboratories served as *S*s. *S*s were assigned at random to three groups of 15 each, with the stipulation that each group have the same proportion of boys to girls.

Apparatus

The apparatus was mounted in a window, 33 in. from the floor, which separated an observation booth from an experimental room. Only the face of the apparatus was visible to *S*, who responded from the experimental room, while *E* recorded responses and operated the necessary controls from the booth.

The face of the apparatus consisted of a black wooden panel, 24 by 16 in. Extending 2½ in. out from the panel were two plexiglass windows 5 in. square, behind which were placed two identical orange stimulus cards. Mounted between the two windows, which were 6¾ in. apart, was a vertically placed 11-in. transparent plastic tube filled with marbles and a marble dispenser which fed into a receiving cup. A rod extending across the front of the apparatus held a curtain which could be drawn to conceal the apparatus from *S*'s view.

Each window was mounted on the weighing platform of a Hanson Utility Scale (Model No. 3000—25 lb. capacity) in which the recording mechanism had been disengaged from the return spring, allowing the scale pointer to remain at the point of maximum push until returned to the zero position by *E*. A microswitch, attached to the spring of each scale, activated a marble release mechanism when a force of ½ lb. or more was delivered to the correct window. A single push against either window cut off the circuit until it was manually reset by *E*; thus, *S* was prevented from obtaining more than one marble per trial or from obtaining a marble by pushing the correct window after pushing the incorrect one.

Additional materials included a wooden stand on which *S* stood during the experiment and a 32-in. marble board with 11 holes. A variety of dime-store toys served as rewards for filling the marble board.

Procedure

S was brought into the experimental room and the apparatus was introduced as a game in which he would be given an opportunity to win a prize. He was told that each time the curtain opened he was to push one of the windows and that, if he pushed the correct one, a marble would drop into the cup, but if he pushed the incorrect one, the marble would not be delivered and he would have to wait until the curtain opened the next time before pushing again.

E then left the room and started pretraining by opening the curtain from the control booth. As soon as *S* pushed one of the windows, the curtain was closed and the window to which the response had been made and the amplitude of that response were recorded. The reopening of the curtain served as a signal for the beginning of the next trial. *S*s in one group were pretrained until they had made 4 correct responses; a second group, 17 correct responses; and the third group, 22 correct responses.[2]

Following pretraining, *E* returned to the experimental room and placed a board with 11 holes on the ledge below the apparatus, directly in front of *S*. *S* was told that the way to win a prize was to earn enough additional marbles to fill all the holes in the marble board. Before training was resumed, *S* selected the toy he wished to win from a display of dime-store toys.

When *S* had filled a predetermined number of holes in the board, marble reinforcements were withheld for responses to either window for a period of seven trials. Following these seven trials, responses to the window which had been incorrect prior to the failure of reward were reinforced until *S* had filled the remaining holes in the board and had won his prize.

One treatment group, 13-N, was introduced to nonreward near the end of the marble board after 13 reinforced training trials. A second group, 26-N, also experienced nonreward near the end of the board but after 26 reinforced training trials. The third group, 26-F, received the larger number of reinforced training trials and was nonrewarded far from the goal. Table 1 summarizes the conditions under which nonreinforcement was introduced for the three groups. It will be noted that group 26-N received twice as many reinforcements prior to nonreward as group 13-N, with distance from the goal held constant. Groups 26-N and 26-F, on the other hand, received an equal number of reinforcements but experienced nonreinforcement at different distances from the goal.

Within each of the three groups, eight *S*s were reinforced for responding to the left window and seven *S*s for responding to the right window. In addition to the 45 *S*s employed, one *S* was omitted from group 13-N due

---

[2] Varying the number of correct responses made during pretraining made it possible to employ the same 11-hole marble board for all three groups.

TABLE 1

CONDITIONS UNDER WHICH NONREWARD WAS INTRODUCED FOR THE THREE TREATMENT GROUPS

	LEVEL OF REINFORCED TRAINING			
Group	*Correct Responses During Pretraining*	*Correct Responses During Marble Board Training*	*Total Number of Correct Responses*	*Marbles Needed to Reach Goal*
13-N	4	9	13	2
26-N	17	9	26	2
26-F	22	4	26	7

to an experimenter error in procedure. One *S* in group 26-N and two in group 26-F were eliminated because of apparatus failure. Three more *S*s, one in group 26-N, and two in group 26-F, were eliminated during pretraining when they persisted in responding either alternately or randomly to the incorrect window.

RESULTS

*S*s made few errors in learning to respond to the correct stimulus window. Groups 26-N and 26-F, in reaching a criterion of 26 correct responses, made 2.46 and 3.60 mean errors, respectively. These means did not differ significantly. Group 13-N, in making 13 correct responses, made .80 mean errors. All errors occurred early in pretraining; thus, all *S*s were responding consistently to the reinforced window prior to the omission of reward.

In order to provide a stable estimate of changes in amplitude occurring as a consequence of nonreinforcement, two scores were obtained for each *S*, the mean force exerted on the four trials preceding the trial on which nonreward was introduced (prescore), and the mean force exerted on the four trials following the first nonreward trial (postscore).[3] Mean pre- and postscores for each group are graphically presented in Figure 1. Inasmuch as there was a tendency for the distributions of these scores to be positively skewed with positively correlated means and variances, the statistical analyses reported below were performed upon a square root transformation of these scores.[4] Means and SDs of the transformed scores are presented in Table 2.

---

[3] Amplitudes recorded for the first nonreinforced trial were not included in the postscore since some *S*s seemed aware that the marble was not being delivered before they had completed the push (making this, effectively, a nonreward trial), while other *S*s did not appear aware that the marble was being withheld until the push had been completed (making this a reward trial).

[4] Employing the 5 per cent level as the coefficient of risk, *F* tests carried out on the nontransformed data led to the same conclusions regarding the experimental hypotheses as those given below.

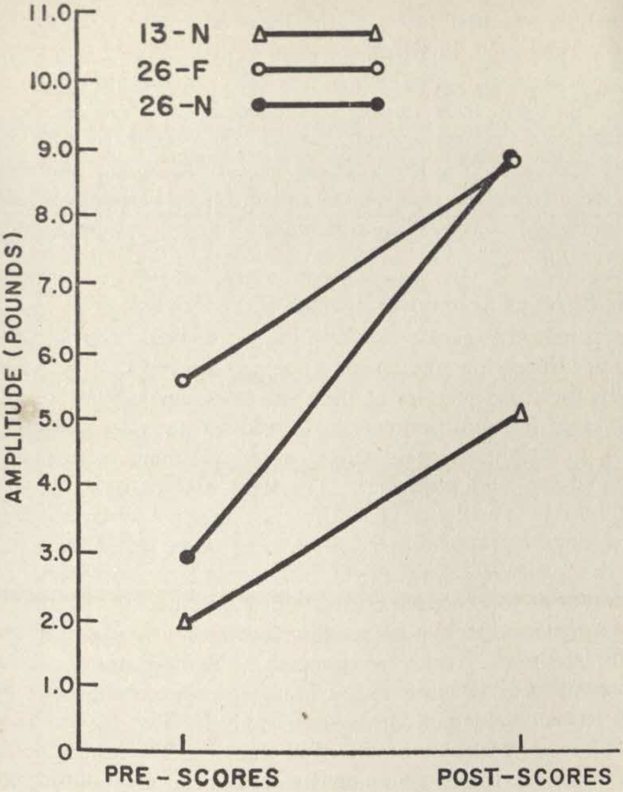

FIGURE 1—Mean force exerted on the four trials prior to the first nonreward trial and mean force exerted on the four trials following the first nonreward trial.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF TRANSFORMED PRE- AND POSTSCORES FOR THE THREE TREATMENT GROUPS

	PRESCORES		POSTSCORES	
Group	*Mean*	*SD*	*Mean*	*SD*
13-N	1.38	.28	2.22	.34
26-N	1.65	.43	2.93	.54
26-F	2.23	.78	2.85	.84

Performance Prior to Nonreward

An analysis was first made of the transformed prescores. A factorial analysis was conducted, in which the main effects were treatment and position of correct window. Only the treatment effect was significant ($p < .001$), indicating that there were reliable differences in the force exerted by the three groups prior to the introduction of nonreward. Separate analyses, comparing groups 26-N and 13-N and groups 26-N and 26-F, indicated that the force exerted by groups 26-N and 13-N did not differ significantly, but that group 26-F was pushing significantly harder than group 26-N prior to nonreward.

Change in Performance as a Consequence of Nonreward

It is apparent in Figure 1 that each group showed a substantial increase in amplitude following the introduction of nonreward. To investigate, statistically, the characteristics of these increases, an analysis was made of the transformed pre- and postscores. A Lindquist Type III analysis (5) was conducted, in which the main effects were treatment, position of correct window, and pre- and postscores.[5] The treatment and the pre-post main effects and the pre-post by treatment interaction were all highly significant, indicating that the groups differed in mean force exerted, that there was an increase in amplitude following the introduction of nonreward, and that the magnitude of this increase varied with experimental treatment.

To test the specific hypotheses that increases in amplitude would be greater for group 26-N than for group 13-N and greater for group 26-N than for group 26-F, separate Type III analyses were conducted, employing the error terms obtained in the over-all analysis. The analysis comparing groups 26-N and 13-N is summarized in Table 3. The results indicate that more force was exerted by group 26-N with both groups showing a significant increase in force following the omission of reward. Of primary interest, however, is the pre-post by treatment interaction which indicates that group 26-N showed a significantly greater increment in amplitude than group 13-N. The significant interaction between pre-post scores, treatment, and position reflects greater differential increases in force, as a function of treatment, for *S*s responding to the right stimulus window.

The analysis comparing groups 26-N and 26-F is also summarized in Table 3. The results show a significant increase in amplitude following the omission of reward for both groups, with a significantly greater increase occurring for group 26-N.

An examination of the pre- and postscores for all 45 *S*s showed that 44 of them showed an increase in response amplitude following the omission of reward. The 45th *S* showed no change.

---

[5] Since the major hypotheses of this experiment concern the pre-post by treatment interactions, and since the proper test of this effect assumes homogeneity of the pre-post by subject interactions, Bartlett's test for homogeneity was applied to the variances of the differences between the transformed pre- and postscores for each treatment group. The resultant χ^2 was 3.86 ($p > .10$).

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE OF TRANSFORMED PRE- AND POST-SCORES FOR GROUPS 26-N AND 13-N AND FOR GROUPS 26-N AND 26-F

		GROUP 26-N VS. GROUP 13-N			GROUP 26-N VS. GROUP 26-F		
Source	*df*	*MS*	*F*	*p*	*MS*	*F*	*p*
Subjects	29						
Treatment (T)	1	3.67	6.12	<.025	.92	1.53	>.20
Position (P)	1	.35			1.97	3.28	>.10
T × P	1	.76	1.27	>.20	.01		
error (b)*	36	.60			.60		
Within Subjects	30						
Pre-Post (PP)	1	16.87	153.36	<.001	13.71	124.64	<.001
PP × T	1	.75	6.82	<.025	1.61	14.64	<.001
PP × P	1	0.00			.11		
PP × P × T	1	.51	4.64	<.05	.13	1.18	>.20
error (w)*	36	.11			.11		

* Error terms are those obtained in over-all analysis.

Performance as a Function of Successive Nonreward Trials

Although the largest increments in amplitude occurred on the trial following the introduction of nonreward, amplitude measures for all seven nonreward trials suggest that amplitude continued to increase with successive nonreinforcements. The statistical significance of this trend was tested by a Lindquist Type I analysis (5). A highly significant trial effect ($p <$ .001) indicated that force did continue to increase with successive nonreward trials.

Persistence of the Originally Correct Response

Since *Ss* continued to respond until they had won enough marbles by pushing the previously incorrect window, it was possible to obtain a measure of persistence of the response to the originally correct window. The number of responses made following the introduction of nonreward, before *Ss* responded to the window which had been incorrect during training, constituted the persistence measure. The mean persistence score for group 26-N was 7.20, for group 26-F, 4.00, and for group 13-N, 6.20. The distributions of these scores were extremely variable, however, and differences among the groups were not statistically significant.

Discussion

The results indicate that *S*s given a relatively large number of reinforcements and nonrewarded near the goal showed a greater increase in amplitude than (a) *S*s given fewer reinforcements and nonrewarded near the goal and (b) *S*s given a large number of reinforcements and nonrewarded far from the goal. It should be recalled, however, that group 26-F was pushing harder prior to nonreward than group 26-N. This suggests the possibility that the differential increase in amplitude occurring for the two groups given the same number of reinforcements but nonrewarded at different distances from the goal was a consequence of a ceiling effect imposed upon the performance of group 26-F. That is, some *S*s in group 26-F may have been responding near maximum prior to nonreward and were thus physically limited in the amount of increment that could accrue as a consequence of nonreinforcement.

The data offer two lines of evidence relevant to such an interpretation. First, if a ceiling effect did act to limit the increment in amplitude occurring for group 26-F, one might expect this group to show a negative correlation between prescores and amount of increase in amplitude following nonreward. The correlation between transformed prescores and amount of increase (obtained by subtracting transformed prescores from transformed postscores) was —.01. There was no significant tendency, therefore, for high amplitude scores prior to nonreward to be associated with small increases in amplitude following nonreward. Secondly, it was possible to match seven *S*s in group 26-N within ½ lb. of the prescores of seven *S*s in group 26-F. Amount of increase in amplitude was compared for these *S*s. A t value of 4.56 (6 df, $p < .01$) indicated a significantly greater increase in amplitude for *S*s in group 26-N than for *S*s in group 26-F. This evidence appears to justify the conclusion that the differential change in amplitude occurring for groups 26-N and 26-F was a consequence of experimental manipulation of goal proximity, rather than a consequence of differences in amplitude level prior to the introduction of nonreward.

The tendency for amplitude to increase with successive nonreward trials suggests that the effect of nonreward upon performance was, up to some point, cumulative. It is of interest to note in this connection that *S*s who persisted for some trials in responding to the originally correct window (the nonreinforced one after the failure of reward) tended to maintain a high level of amplitude, if not increase it, for several trials following the seven-trial nonreinforcement period.

The fact that 44 of 45 *S*s showed an increase in amplitude following the omission of reward accords with the hypothesis (2) that omission of a customary reward gives rise to an emotional state (frustration) which increases drive and hence level of performance. The finding that such increases in performance are positively related to number of reinforcements received prior to nonreward and the proximity of the point of nonreward from the goal offers support for the suggestion (9) that degree of frustration is some function of the amount of "reward expectation" which has been developed.

SUMMARY

Three groups of preschool children learned to push the correct one of two stimulus windows for a marble reinforcement. After a predetermined number of correct responses had been made, reinforcement was omitted for seven trials. After these nonreward trials, responses to the window which had been incorrect during training were rewarded until *S* had obtained enough marbles to finish filling a marble board, entitling him to a toy reward.

Mean force exerted on the four trials prior to the introduction of nonreward was compared with the mean force exerted on the four trials following the introduction of nonreward in order to determine the effect of reward failure upon increases in amplitude of response.

Results showed that, with goal distance controlled, a group receiving a relatively high number of reinforcements prior to nonreward showed a significantly greater increase in amplitude following nonreinforcement than a group receiving fewer reinforcements. Furthermore, with number of reinforcements controlled, a group introduced to nonreward close to the completion of the marble board showed a significantly greater increase in amplitude than a group experiencing nonreward at a point farther removed from the goal. Forty-four of the total 45 *S*s showed an increase in amplitude following the omission of reinforcement.

It was suggested that the experimental variables subjected to manipulation in this investigation may have differentially influenced the degree of frustration resulting from the omission of a customary reward.

REFERENCES

1. AMSEL, A., & ROUSSEL, J. Motivation properties of frustration: effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.,* 1952, 43, 363-368.
2. BROWN, J. S., & FARBER, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.,* 1951, 48, 465-495.
3. HANER, C. F., & BROWN, P. A. Clarification of the instigation to action concept in the frustration-aggression hypothesis. *J. abnorm. soc. Psychol.,* 1955, 51, 204-206.
4. HILGARD, E., & MARQUIS, D. *Conditioning and learning.* Appleton-Century-Crofts, 1940.
5. LINDQUIST, E. *Design and analysis of experiments in psychology and education.* Houghton Mifflin, 1953.

6. Marzocco, F. N. Frustration effect as a function of drive level, habit strength, and distribution of trials during extinction. Unpublished doctoral dissertation, State Univer. of Iowa, 1951.

7. Roussel, J. Frustration effect as a function of repeated nonreinforcements and as a function of the consistency of reinforcement prior to the introduction of nonreinforcement. Unpublished master's thesis, Tulane Univer., 1952.

8. Skinner, B. F. *Behavior of organisms*. Appleton-Century-Crofts, 1938.

9. Wilson, W., Weiss, E. J., & Amsel, A. Two tests of the Sheffield hypothesis concerning resistance to extinction, partial reinforcement, and distribution of practice. *J. exp. Psychol.*, 1955, 50, 51-60.

RELATIVE EFFECTIVENESS OF DIFFERENT REINFORCEMENT COMBINATIONS ON CONCEPT LEARNING OF CHILDREN AT TWO DEVELOPMENTAL LEVELS

William J. Meyer* *and* Stanley B. Seidman
University of Pittsburgh

In a recent paper (7) the writers attempted to demonstrate that under certain conditions developmental level is an important variable in the determination of the effectiveness of various types of reinforcements. Reasoning that under relatively ambiguous learning conditions older *S*s would be sensitive to a larger variety of reinforcing cues, an experiment was designed in which the *S*s received no prior instructions concerning the nature or meaning of the reinforcements employed. The reinforcement combinations were as follows: *Right-Wrong* (RW), *Right-Nothing* (RN), *Nothing-Wrong* (NW), and *Buzzer-Nothing* (BN). It was assumed that the presence of the *nothing* condition in three of the combinations would provide sufficient ambiguity to permit differential interpretation of the examiner's behavior and that such differences might be related to developmental level. It was further anticipated that the presentation of the buzzer without specifying its meaning, i.e., the buzzer could either signify a correct or an incorrect response, would also provide an ambiguous reinforcement. Since little is known about the nature of extraneous reinforcement cues, no formal predictions concerning the experiment were attempted.

Experimental verification was attempted by randomly placing 20 youngsters in each of the four treatment groups and at each of two developmental levels, four- and five-year-olds and eight- and nine-year-olds, affording a total of 160 *S*s. The task consisted of a simple discrimination problem, learning to respond to the larger of two blocks. An analysis of variance of the acquisition data revealed a significant trials × age interaction and a significant trials × age × treatments interaction. On the basis of these results it was concluded that the effectiveness of the four reinforcement combinations was related to age and that rate of acquisition was a function of age and treatments. Multiple comparison tests failed to yield significant differences, but inspection of the learning curves suggested that the NW combination was more effective for the older as compared with the younger *S*s whereas the RW and BN combinations were more effective for the younger than the older *S*s. Since the multiple comparison tests were not

---

[*] Department of Psychology, University of Pittsburgh, Pittsburgh 13.

Child Develpm., 1961, 32, 117-127.

statistically significant, it was decided to replicate part of the original study by comparing the RW-NW combinations in one experiment and the NW-BN combinations in a second experiment.

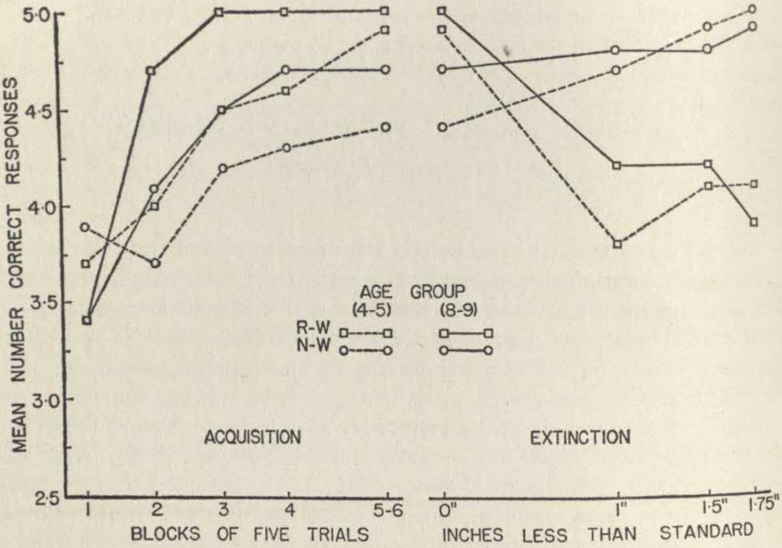

FIGURE 1—Mean number of correct responses on acquisition and extinction series for RW and NW treatments.

EXPERIMENT I

Procedure

The first experiment compared the RW-NW combinations. A total of 40 *S*s, 20 four- and five-year-olds and 20 eight- and nine-year-olds attending a suburban public school, constituted the sample.[1] The *S*s were from an average socioeconomic environment and were representative of an average intellectual population.

Ten *S*s were randomly assigned to each of the four experimental groups comprised of two treatments and two developmental levels. The *S*s in the RW group were told "right" following a correct response and "wrong" following an incorrect response. The NW group was told *nothing* following a correct response and "wrong" following an incorrect response.

The stimuli for the learning problem were identical to those employed in the earlier study. They comprised nine pairs of three-dimensional geometric blocks in the shape of cylinders, triangles, and rectangles. The color (red) and heights (2 in.) of all blocks were identical; only the basal surface area varied from 1 to 2 in. These stimuli comprised the acquisition series. For the extinction trials, the same basal area relationships (2 : 1) were maintained, but the heights of the stimuli were reduced by 1 in., 1.5 in., and 1.75 in. These stimuli constituted the extinction series. The stimuli were presented in pairs; the position of the correct choice was randomized with the larger basal area always being the correct block. Additional apparatus consisted of a portable box that had a board attached to it which served as a blind.

---

[1] The writers wish to express their appreciation to Mr. Allen Harris, supervising principal of the Eastmont School, for his cooperation in making *S*s available for this study.

The *S*s were tested individually. Each *S* received the same instructions: "I am going to show you two blocks at a time. Each time you must choose, by pointing with your finger, the block that you think is the correct one. Remember that one block is always the right one to choose and one block is always the wrong one to choose." After reaching a criterion of 9 out of 10 correct responses or a maximum of 30 trials, whichever came first, each *S* was immediately started on the extinction series of five trials for each of the stimulus conditions. During this series no reinforcements were given.

Results

In the previous study it was reported that little learning occurred after the twentieth trial, hence, only the first 20 trials were included in the statistical analysis. A similar result was found in the present study and was handled in the same way. The data were organized into four blocks of five trials and are presented graphically in Figure 1. Table 1 shows the mean number of correct responses made by the two experimental groups on the 20 trials.

TABLE 1

MEANS AND SIGMAS OF NUMBER OF CORRECT RESPONSES ON ACQUISITION SERIES

	AGE 4 TO 5		AGE 8 TO 9	
Treatment	*Mean*	*Sigma*	*Mean*	*Sigma*
NW	16.10	3.06	16.70	3.26
RW	16.80	2.92	18.10	4.24

A summary of the analysis of variance for repeated measures on the same *S*s, as described by Edwards (6), for the acquisition data is presented in Table 2. The Bartlett test was statistically nonsignificant. The following conclusions are warranted: (a) effects attributable to age, treatments, and the treatments × age interaction are not statistically significant; (b) the between trials effect is statistically significant, i.e., the learning curves are not flat; (c) the trials × treatments, trials × age, and the trials × age × treatments interactions fail to reach an acceptable level of significance. The

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF ACQUISITION DATA FOR RW-NW TREATMENTS

Source	*df*	*Mean Square*	*F*
Treatments (T)	1	2.25	1.87
Age (A)	1	2.75	2.29
T × A	1	.31	
Error A	36	1.20	
Trials	3	9.15	12.20**
Trials × T	3	1.86	2.48
Trials × A	3	.63	
Trials × T × A	3	.03	
Error B	108	.75	

** $p < .01$.

results of the foregoing analysis permit the conclusion that both the RW and NW reinforcement combinations are equally effective at both levels of development.

The data on mean performance of each experimental group on the three sets of stimuli constituting the extinction series are presented in Table 3 and the right side of Figure 1. The analysis of variance relevant to these data is presented in Table 4. This analysis indicates that over all the stimulus pairs there is a statistically significant difference in which the NW treatment is superior to the RW. The significant stimuli effect can be interpreted as meaning that the number of correct responses on each set of stimuli are significantly different. The significant stimuli × treatments interaction indicates that the slopes of the extinction curves are different. Inspection of Figure 1 shows that the NW group tends to improve in performance somewhat whereas the RW group declines.

TABLE 3

MEANS AND SIGMAS OF NUMBER OF CORRECT RESPONSES ON EXTINCTION SERIES

Stimulus Height (in. less than standard)	TREATMENT N-W				TREATMENT R-W			
	AGE 4 TO 5		AGE 8 TO 9		AGE 4 TO 5		AGE 8 TO 9	
	Mean	*Sigma*	*Mean*	*Sigma*	*Mean*	*Sigma*	*Mean*	*Sigma*
0	4.40	.96	4.70	.94	4.90	.32	5.00	0.00
1	4.70	.67	4.80	.63	3.80	1.13	4.20	1.13
1.50	4.90	.32	4.80	.63	4.10	.87	4.20	1.61
1.75	5.00	.00	4.90	.32	4.10	.87	3.90	1.65

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE OF EXTINCTION-GENERALIZATION DATA FOR RW-NW TREATMENTS

Source	*df*	*Mean Square*	*F*
Treatments (T)	1	10.00	5.13*
Ages (A)	1	.12	
T × A	1	.03	
Error A	36	1.95	
Stimuli (S)	3	4.20	10.00**
S × T	3	3.72	8.85**
S × A	3	.34	
S × T × A	3	.14	
Error B	108	.42	

* $p < .05$.
** $p < .01$.

On the basis of our previous research it was expected that the NW treatment would be more effective with the older *S*s whereas the RW treatment would be more effective with the younger *S*s. The statistically nonsignificant treatments × age interaction clearly shows that these pairs of reinforcers are equally effective with both age groups. Though the nonsignificant treatments effect precludes any definitive statements concerning the reinforcement combinations, it should be noted that the RW condition produced the most rapid learning amongst the older *S*s whereas for the same age group in the earlier study the NW treatment produced the most rapid learning. We shall offer an hypothesis later in this paper in an effort to explain this discrepancy.

In terms of the extinction data, the earlier study suggested that the NW group would show little if any decrement in performance whereas the RW *S*s would demonstrate the typical extinction curve. The present data clearly support this expectation.

The results of the acquisition phase of this experiment are in agreement with the findings reported by Buss *et al.* (4) and Buss and Buss (5). These investigators, employing student nurses and psychiatric patients as *S*s, found little difference in learning between their NW and RW groups. They interpreted their results as indicating that *right* is a much weaker reinforcer than *wrong* and that the critical component of the combinations is *wrong*. Our results are consistent with this notion and indicate that the differential strength of *right* and *wrong* exists at least as early as the fifth year of life.

Inconsistent with the Buss study, but consistent with Buchwald (3), our NW group performed better during extinction. In our earlier paper, wherein a RN group was included, the foregoing data were interpreted as meaning that *nothing* when paired with *wrong* acquires positive reinforcing value but when paired with *right* acquires negative reinforcing value. The RW treatment, which obviously does not give the *S*s any cues concerning *nothing*, presumably should not affect its reinforcement value. Both Buss (4) and Buchwald (2, 3) found no differences between NW and RW on extinction and so support the foregoing assertions. The findings of the present study are clearly not in accord with these investigators. In interpreting our data, we take the position that whenever *right* is explicit in a pair of reinforcers, a subsequent failure of this response to occur has a negatively reinforcing effect. A similar effect is not derived from the *wrong* condition because, except for the first 10 trials or so, the *S*s are essentially functioning as an RN group wherein *nothing* acquires negative reinforcement value.

EXPERIMENT II

Procedure

The second experiment compared the NW and BN combinations. A total of 40 *S*s, 20 four- and five-year-olds and 20 eight- and nine-year-olds attending a suburban public school, constituted the sample. The *S*s were from the same school as those in Experiment I.[2] The procedures and stimulus materials were identical to those in Experiment I with the exception that the RW group was replaced by the Buzzer-Nothing (BN) group. This latter group heard a buzzer (which was secured to the board and out of sight of the *S*s) every time they made a correct response and *nothing* for an incorrect response. No explanation of the meaning of the buzzer was given.

TABLE 5

MEANS AND SIGMAS OF NUMBER OF CORRECT RESPONSES ON ACQUISITION SERIES

	AGE 4 TO 5		AGE 8 TO 9	
Treatment	*Mean*	*Sigma*	*Mean*	*Sigma*
NW	16.10	3.06	16.70	3.26
BN	13.60	5.31	13.80	3.02

Results

The relevant data for Experiment II are reported in Table 5 and Figure 2. The data were treated as before, and again the Bartlett tests were not statistically significant. The analysis of variance presented in Table 6

---

[2] The NW *S*s for the second experiment are the same as those used in Experiment I.

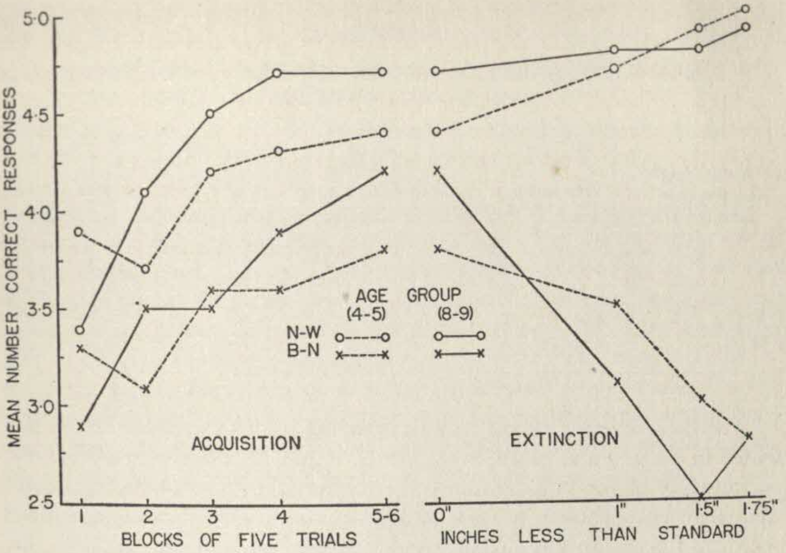

FIGURE 2—Mean number of correct responses on acquisition and extinction series for NW and BN treatments.

permits the following conclusions: (a) there is a significant difference between the two treatments in which the NW group is superior to the BN group; (b) both age groups perform similarly under the two treatments; (c) the nonsignificant treatments × age interaction indicates that the NW treatment is superior for both age groups.

TABLE 6

SUMMARY OF ANALYSIS OF VARIANCE OF ACQUISITION DATA FOR NW-BN TREATMENTS

Source	*df*	*Mean Square*	*F*
Treatments (T)	1	18.23	5.09*
Ages (A)	1	.40	
T × A	1	.10	
Error A	36	3.58	
Trials	3	4.58	7.05**
Trials × T	3	.14	
Trials × A	3	1.51	2.32
Trials × A × T	3	.12	
Error B	108	.65	

* $p < .05$.
** $p < .01$.

TABLE 7

MEANS AND SIGMAS OF NUMBER OF CORRECT RESPONSES ON EXTINCTION SERIES

Stimulus Height (in. less than standard)	TREATMENT N-W				TREATMENT B-N			
	AGE 4 TO 5		AGE 8 TO 9		AGE 4 TO 5		AGE 8 TO 9	
	Mean	*Sigma*	*Mean*	*Sigma*	*Mean*	*Sigma*	*Mean*	*Sigma*
0	4.40	.96	4.70	.94	3.80	1.31	4.20	1.31
1	4.70	.67	4.80	.63	3.50	1.71	3.10	1.28
1.50	4.90	.32	4.80	.63	3.00	1.69	2.50	1.26
1.75	5.00	.00	4.90	.32	2.80	1.54	2.80	1.47

The data relevant to the extinction series for this experiment are presented in Table 7 and on the right side of Figure 2. As was anticipated, the performance of the NW groups is clearly superior. The highly significant treatments effect shown in Table 8 indicates that the NW group in contrast with the BN group gave reliably more correct responses. In addition, the significant stimuli × treatments interaction shows that the slopes of these curves are different with the BN *S*s declining sharply in performance after the discontinuation of reinforcement.

The results of our first study suggested that the BN treatment would be more effective with the younger *S*s and the NW treatment would be more effective with the older Ss. The nonsignificant age × treatments interaction indicates that each age group responds similarly to the two treatments and therefore fails to support the earlier results. In this study the BN *S*s of both

TABLE 8

SUMMARY OF ANALYSIS OF VARIANCE OF EXTINCTION-GENERALIZATION DATA FOR NW-BN TREATMENTS

Source	*df*	*Mean Square*	*F*
Treatments (T)	1	97.66	27.66**
Ages (A)	1	.06	
T × A	1	.30	
Error A	36	3.53	
Stimuli (S)	3	1.75	3.44*
S × T	3	5.57	10.71**
S × A	3	.77	1.28
S × T × A	3	.26	
Error B	108	.52	

* $p < .05$.
** $p < .01$.

age groups performed in a manner similar to the Right-Nothing group included in the original paper and suggests that the buzzer (when undefined) is no more effective than saying "right." The extinction data clearly support the previous findings and would not seem to require further comment here.

Discussion

The failure to replicate the age × treatments interaction requires examination. Since the experimental procedures were identical in every detail throughout the entire series of studies, it does not seem likely that this area of examination would prove fruitful. One possible source of variance perhaps exists in the samples employed in both studies. The *S*s in the earlier study were drawn primarily from a private university elementary school in which the general intellectual level is high (mean IQ = 124) and wherein considerable emphasis is placed on academic achievement and competition. In contrast, the second sample was drawn entirely from a public school wherein the general intellectual level was lower and where the school placed less emphasis upon academic competition. One might speculate that differences in achievement motivation were responsible for the divergent findings. However, we believe that until further data can be brought to bear on this issue the most plausible explanation of the failure to replicate is a statistical consideration. It seems reasonable that the significant interaction resulted from some slight differences in treatments effectiveness which when pooled for eight groups (four treatments and two developmental levels) achieved a magnitude sufficient for statistical significance and resulted in a Type I error.

Summarizing the results of the present study and the earlier report, it can be concluded that for both preschool age children and primary grade children verbal reinforcement combinations which include *wrong* as a component are more effective than just positive reinforcement (RN and BN). At both developmental levels *wrong* appears to be a stronger negative reinforcer than *right* is a positive reinforcer. Upon the discontinuation of all reinforcement it can be concluded that the combination which during acquisition does not include a positive reinforcer (NW) produces the least amount of extinction. The combination which during acquisition includes a positive reinforcer but not a negative reinforcer (RN, BN) produces the greatest amount of extinction. The combination including both a positive and a negative reinforcer (RW), which presumably permits the original strength of *nothing* to operate, leads to extinction which is greater than NW but not as great as RN or BN. The foregoing generalizations were found at both developmental levels.

The results of this and the earlier study have been interpreted from the point of view that *right* and *wrong* are at opposite ends of the reinforcement continuum with *nothing*, assumed to have little or no initial reinforcement value, placed closer to the weaker reinforcer of *right*. We have further taken the position that *nothing,* when paired with *wrong* or *right,* acquires, respectively, positive or negative reinforcement value. However, the writers are not asserting that our position is the only one which explains our data. It can be argued, for example, that the extinction data result merely from the extreme change in conditions from acquisition to extinction for the RW, RN, and BN groups whereas the extinction condition for the NW group remains very similar to the acquisition condition. Though such an explanation makes sense, we would assert that such a position limits the possibilities for further research. By maintaining the theoretical properties ascribed to the *nothing* condition, several definitive experiments are suggested [*see* Buchwald (2), for an example]. These experiments may fail to support our position, but, at least, they will have the value of increasing our understanding of reinforcement effectiveness amongst children.

SUMMARY

In a previous experiment involving four reinforcement combinations (Right-Wrong, Right-Nothing, Nothing-Wrong, and Buzzer-Nothing) and two developmental levels (four- and five-year-olds and eight- and nine-year-olds), a significant age × treatments interaction was reported. Since multiple comparison tests failed to yield significant results, it was decided to replicate the study using those treatments which seemed to contribute most to the interaction terms (RW, NW, BN). Two separate experiments were conducted using age groups and procedures identical to the original study. Comparison of the RW-NW combinations and the NW-BN combinations constituted the first and second experiments, respectively. The results of the experiments failed to replicate the age × treatment interaction. No difference was found between the RW-NW treatments though the NW group was reliably superior to the BN group. The extinction data replicated in entirety. Reasons for the lack of agreement for the acquisition results were examined. On the basis of the two studies the conclusion was reached that *wrong* is a stronger reinforcer than *right* and that this difference appears as early as five years of age. It was also concluded that regardless of developmental level *nothing* when paired with *wrong* acquires positive reinforcement value and serves to maintain performance at a pre-extinction rate. When *nothing* is paired with *right,* it acquires negative reinforcement value and serves to extinguish behavior.

REFERENCES

1. BRACKBILL, Y., & O'HARA, J. The relative effectiveness of reward and punishment for discrimination learning in children. *J. comp. physiol. Psychol.,* 1958, 6, 747-751.

2. Buchwald, A. M. Experimental alterations in the effectiveness of verbal reinforcement combinations. *J. exp. Psychol.,* 1959, 57, 351-361.

3. Buchwald, A. M. Extinction after acquisition under different verbal reinforcement combinations. *J. exp. Psychol.,* 1959, 57, 43-48.

4. Buss, A. H., Braden, W., Orgel, A., & Buss, E. Acquisition and extinction with different verbal reinforcement combinations. *J. exp. Psychol.,* 1956, 52, 288-295.

5. Buss, A. H., & Buss, E. The effect of verbal reinforcement combinations on conceptual learning. *J. exp. Psychol.,* 1956, 52, 283-287.

6. Edwards, A. L. *Experimental design in psychological research.* Rinehart, 1950.

7. Meyer, W. J., & Seidman, S. B. Age differences in the effectiveness of different reinforcement combinations on the acquisition and extinction of a simple concept learning problem. *Child Develpm.,* 1960, 31, 419-429.

NORMATIVE DATA ON THE CHILDREN'S MANIFEST ANXIETY SCALE AT THE RURAL THIRD GRADE LEVEL [1]

HAROLD D. HOLLOWAY*
University of Tennessee

The children's form of the manifest anxiety scale (CMAS) developed recently by Castaneda, McCandless, and Palermo (1) has become a valuable research instrument at the fourth, fifth, and sixth grade levels. The predictive merit and reliability of the CMAS using these mid-grade subjects (*S*s) may be confirmed by referring to the series of articles, given in the references, written jointly by the three above authors and singly by Palermo (7). Holloway (3) presented evidence that the CMAS could also be used reliably (retest $r = .83$) at the rural third grade level.

The purpose of the present study was to furnish additional information concerning the CMAS in the form of normative statistics at the rural third grade level—an essential complement to any newly devised instrument.

METHOD

Subjects

The sample was actually a byproduct of a larger study (hereafter "main study") concerning relations between parental attitudes toward childrearing and the CMAS in Tennessee. In order to obtain CMAS protocols designated by the sampling design of the main study,[2] it was necessary to test all pupils

---

* Department of Child Development and Family Relationships, University of Tennessee, Knoxville.

[1] The study was sponsored jointly by the Agricultural Experiment Station and the College of Home Economics, Department of Child Development and Family Relationships, of the University of Tennessee. The author wishes to express his gratitude to the public school officials and teachers who participated and granted permission to conduct the study. The author also acknowledges his gratitude to Dr. Mary E. Keister, University of Tennessee, and to Dr. A. L. Finkner, Department of Experimental Statistics, North Carolina State College, who participated as consultants.

[2] The design for the main study was a one-stage stratified random sample, the strata being the three principal geographical regions of Tennessee—East, Middle, and West. The basic sampling unit was a cluster of approximately 15 third graders from a given rural school.

Child Develpm., 1961, 32, 129-134.

in a given third even though in many instances but a few of the pupils in the room were slated for use in the main study. The procedure resulted in additional usable CMAS protocols. The total number obtained was 524 with approximately 200 of them to be used in the main study. The sample of the present study was constructed from among the 524 protocols simply by equalizing, for analytical reasons, the number of boys and girls by a random procedure within each of three main geographical strata of Tennessee, namely, East, Middle, and West. The resultant number of *Ss* was 462 distributed with respect to sex, number, age, and location as shown in Table 1. The sample cannot be described in the strict sense as being a so-called "representative" or "random" one of the state. However, the results to be reported will constitute the best estimate one can make in such a direction at the present time.

TABLE 1

SEX, NUMBER, AGE, AND LOCATION CHARACTERISTICS OF THE SUBJECTS

	SEX AND NUMBER			CA IN YEARS			
Stratum	*Boys*	*Girls*	*Total*	*Mean*	*Range*	*No. of Counties*	*No. of Communities*
East	103	103	206	9.0	7.7–15.2	6	10
Middle	87	87	174	8.8	7.8–12.0	5	5
West	41	41	82	8.9	8.1–11.0	3	3
Combined ..	231	231	462	8.9	7.7–15.2	14	18

The subjects were white, third graders, attending schools in rural areas ("rural" as defined by the 1950 U. S. Census, i.e., open country or places with less than 2500 population—the four main metropolitan counties were eliminated) and attending schools having more than one teacher.

Procedure

The CMAS was administered orally to each classroom by the teacher who followed a set of standard instructions prepared by the present author. The testing program was accomplished during 1957 and 1958.

RESULTS AND DISCUSSION

The obtained A scale (42-item "anxiety" subscale of the CMAS) and L scale (11-item "tendency to falsify responses to A" subscale) scores were entered into separate Stratum × Sex analysis of variance designs.[3] The basic means and SDs of the two subscales are presented in Table 2 combined

---

[3] For a complete description of the CMAS and its A and L subscales, see Castaneda, McCandless, and Palermo (1).

TABLE 2

STRATUM × SEX MEANS AND STANDARD DEVIATIONS FOR A AND L SCALES SEPARATELY

	BOYS		GIRLS		COMBINED SEXES	
Stratum	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
A Scale						
East	22.27	7.28	23.79	7.78	23.03	7.57
Middle	20.05	8.40	20.77	8.05	20.41	8.24
West	19.66	9.58	19.78	9.33	19.72	9.46
Combined Strata	20.97	8.24	21.94	8.35	21.45	8.31
L Scale						
East	5.37	1.59	5.94	1.80	5.66	1.72
Middle	5.28	1.84	5.59	1.75	5.43	1.80
West	5.68	1.69	5.73	1.71	5.71	1.70
Combined Strata	5.39	1.71	5.77	1.77	5.58	1.75

with the total sample statistics. The total sample values will be considered first to be followed by separate presentations of the analyses of variance and the correlations between the A and L scales.

Total Sample Statistics

The A scale mean and SD for the total sample of 462 *S*s are given in the upper-right section of Table 2. The median (21.98), eightieth and twentieth percentiles ($P_{80} = 29.30$ and $P_{20} = 14.04$, respectively) are reproduced in Figure 1 which is an over-all frequency polygon of the A scale scores. In general, the separate sex and stratum distributions were similar to that shown in Figure 1, that is, approximately bell shaped and unimodal. For girls alone the median was 22.67; P_{80} was 29.84, and P_{20} was 14.90. The corresponding values for boys were: 21.40, 28.47, and 13.10.

Inspection of the frequency distributions plotted similarly for the total sample of L scale scores revealed them also to be approximately bell shaped and unimodal. The mean and SD of the 462 L scale scores are given in the lower-right section of Table 2; the median was 5.74; P_{80} was 7.15, and P_{20} was 4.02. The same statistics for girls were: 5.92, 7.32, and 4.24; for boys, 5.56, 6.96, and 3.86.

Analyses of Variance

Summaries of the two Stratum × Sex analyses of variance, one for the A scale and one for the L scale, have been combined and presented in Table 3.

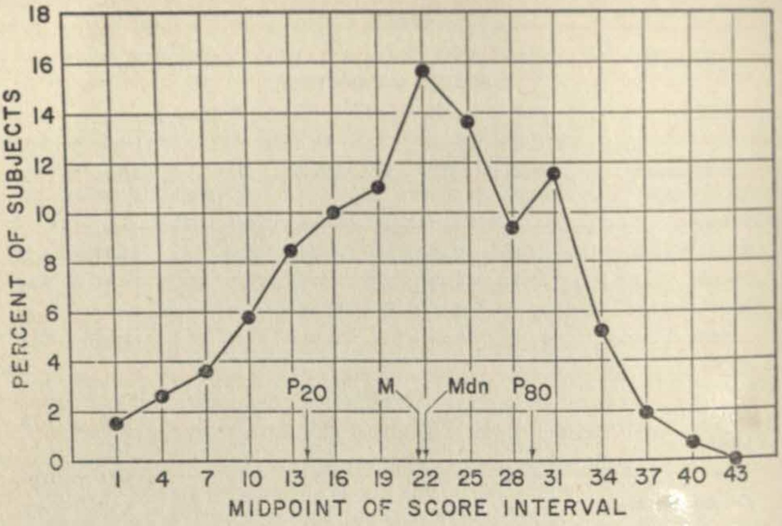

FIGURE 1—Frequency polygon showing percentage of 462 third grade rural children receiving the indicated scores on the A scale of the CMAS.

A scale. It may be observed from the A scale section of Table 3 that the one statistically significant *F* ratio (7.02) indicated an over-all difference among the stratum means. Inspection of these three means in the upper right portion of Table 2 revealed a consistent increase from West (19.72), through Middle (20.41), to East Tennessee (23.03). By definition, the

TABLE 3

STRATUM × SEX ANALYSIS OF VARIANCE SUMMARIES FOR A AND L SCALES SEPARATELY

		A SCALE		L SCALE	
Source of Variation	*df*	*Mean Square*	*F*	*Mean Square*	*F*
Strata	2	474.07	7.02**	3.18	1.04
Sex	1	108.61	1.61	16.76	5.50*
Interaction	2	16.32	< 1.00	2.19	< 1.00
Within-Cells	456	67.53		3.05	
Total	461	69.16		3.08	

NOTE.—Coefficient of risk of $p = .05$ used for all *F* tests.
* $p < .05$.
** $p < .01$.

higher the A scale score, the higher the anxiety. Individual *t* tests[4] between pairs of means showed that the East mean was significantly higher than either the Middle ($t = 3.08$; $p < .01$) or West ($t = 3.08$; $p < .01$) means with no significant difference between the latter two strata ($t < 1.00$).

Generally speaking, the means were higher than those found in similar studies involving fourth, fifth, and sixth graders in different parts of the country. Whether or not such differences can be attributed to variation in grade level and/or geographical location remains unanswered. The trend for the A scale means to increase from West to East was perhaps the most intriguing result, and one can only speculate as to the reasons. The most obvious possibility (or strata correlate) is the fact that population size, like the anxiety means, increases from West to East. Secondly, the terrain ranges from relatively flat farming land in the West, through semimountainous regions in Middle, to mountainous sections in the East. Sociological experts of the state would certainly posit general cultural differences between peoples living in rural mountain communities versus those in rural nonmountain communities. The author would hypothesize that among such cultural differences are such factors as conservatism and restrictiveness in social philosophy and behavior. It would seem that these two characteristics are weighted more heavily in the East than in the West which, on a speculative basis, is perhaps reflected in increasing anxiety from West to East.

L scale. The analysis of variance yielded a significant *F* ratio for the sex factor ($F = 5.50$—Table 3), indicating that the mean for girls of 5.77 was significantly higher than the mean of 5.39 for boys (*see* Table 2). The finding was consistent with the research literature involving the CMAS.

A and L Scale Correlations

Ideally, from the standpoint of measurement, the correlation between the anxiety portion of the CMAS (A scale) and the "tendency to falsify" the anxiety items (L scale) should be zero. Pearson *r*s were computed between A and L for each stratum-sex combination to ascertain an estimate of the relationship. The obtained *r*s ranged between −.23 and .00, none being significantly different from zero using a coefficient of risk of $p = .05$. These correlations were typical of those found in other investigations.

SUMMARY

The purpose of the study was to obtain normative data on the CMAS at the rural third grade level. The CMAS was administered orally by the teachers of 462 third grade rural children, who were equally divided with respect to sex and who ranged in age from 7.7 to 15.2 years with a mean of 8.9 years. The *S*s were drawn from 18 rural communities distributed throughout East, Middle, and West Tennessee. Data were analyzed in terms of frequency distributions, Stratum × Sex analyses of variance, and correlations. The main results were:

---

[4] Within-cells mean square used as a common estimate of error variance.

1. The principal A and L scale frequency distributions were approximately bell shaped and unimodal.

2. The A scale means increased consistently from West to East Tennessee with a significant difference occurring between East and each of the other two strata. Sex differences and their interaction with strata were not significant.

3. The L scale mean for girls was significantly higher than for boys. Stratum differences and the stratum-sex interaction were not significant.

4. Pearson *r*s between the A and L scales were not significantly different from zero.

REFERENCES

1. CASTANEDA, A., MCCANDLESS, B. R., & PALERMO, D. S. The children's form of the manifest anxiety scale. *Child Develpm.,* 1956, 27, 317-326.
2. CASTANEDA, A., PALERMO, D. S., & MCCANDLESS, B. R. Complex learning and performance as a function of anxiety in children and task difficulty. *Child Develpm.,* 1956, 27, 327-332.
3. HOLLOWAY, H. D. Reliability of the children's manifest anxiety scale at the rural third grade level. *J. educ. Psychol.,* 1958, 49, 193-196.
4. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Houghton Mifflin, 1953.
5. MCCANDLESS, B. R., & CASTANEDA, A. Anxiety in children, school achievement and intelligence. *Child Develpm.* 1956, 27, 378-382.
6. MCCANDLESS, B. R., CASTANEDA, A., & PALERMO, D. S. Anxiety in children and social status. *Child Develpm.,* 1956, 27, 385-391.
7. PALERMO, D. S. Racial comparisons and additional normative data on the children's manifest anxiety scale. *Child Develpm.,* 1959, 30, 53-57.
8. PALERMO, D. S., CASTANEDA, A., & MCCANDLESS, B. R. The relationship of anxiety in children to performance in a complex learning task. *Child Develpm.,* 1956, 27, 333-337.

SOCIAL HISTORIES OF SUCCESSFUL AND UNSUCCESSFUL CHILDREN

JOHN W. BAKER* *and* ANNETTE HOLZWORTH[1]
Mental Health Research Institute, Fort Steilacoom, Washington

The social histories of children in a state mental hospital and of successful children in a matched control group will be statistically analyzed and compared in this paper.

Information obtained from social histories has long been used in the diagnosis of children's emotional and social problems as well as in the formulation of treatment plans. In more recent years, the social history has been a part of research studies relating to children's problems. Although isolated items found in social histories have been statistically analyzed, the statistical analysis of complete social histories has been infrequent. Obviously, many items in a social history do not lend themselves to statistical analysis. In addition, the social situation of a family or individual is not accurately or completely portrayed in statistical studies. However, because of the widespread use of the social history as a diagnostic and research instrument, a statistical examination of the social histories of various groups is pertinent. A review of the literature did not reveal any studies that attempted a statistical comparison of the complete social histories of successful school children and of children admitted to a psychiatric hospital, two groups that differ greatly in their social adjustment. It is for these reasons that this study should be of interest to professional people who work with emotionally disturbed children.

METHODOLOGY

Seventy-one children between the ages of 13 and 16 years comprise each of the two groups, making a total of 142 subjects included in this study. The social work unit of the Mental Health Research Institute obtained the social histories of these children between the years 1955 and 1958. The two

---

* Supervisor, Psychiatric Social Services, Division of Mental Health, Department of Institutions, State of Washington, P.O. Box 867, Olympia.

[1] Former Research Assistant, Mental Health Research Institute, Fort Steilacoom, Washington.

Child Develpm., 1961, 32, 135-149.

groups were matched for age and sex. Eighty-four of the subjects were matched with less than two months difference in date of birth, with the largest gap being four and one-half months. The patients were children admitted to Western State Hospital, Fort Steilacoom, Washington during the years of the study. The control group of successful children met the following criteria: (a) they were in the appropriate school grade, and (b) they had a history of good citizenship in school. This control group was chosen at random by selecting subjects at regular intervals from a larger list of successful students in two public schools in Tacoma, Washington. School officials—teachers, guidance department personnel, and principals—selected the original list of students. Therefore, the control group was not just a random sample of children in two public schools but were successful students by school standards. The two schools were chosen from an area thought to be somewhat similar to the patient group in socioeconomic status.

Each group consisted of 35 boys and 36 girls with a total of 70 boys and 72 girls in the study population. In the patient group there were two Negroes and two with mixed racial backgrounds, while the controls were all Caucasian. Although the psychiatric diagnoses of the patient group are not considered in the findings of this paper, they are as follows: schizophrenic—11, schizophrenic with indications of organicity—17, primary neurotic—15, neurotic with a questionable thought disturbance—13, and neurotic with psychopathic manifestations—15. The diagnoses were made by the psychiatric staff of the Mental Health Research Institute.

A psychiatric social worker obtained the social histories from the parents of the control group during an interview lasting approximately two hours in either the parents' home or at the school. The same social history outline was used with the parents of the patient group. When reports from other agencies were available, they were used to supplement the patients' histories. Several different interviewers participated in the study and attempted to obtain as detailed a history as possible in the allotted time.

The outline used in taking the histories covered subjects that are generally used in social histories and perhaps in more detail than usual in a service agency. The parents were asked about the child's prenatal, birth and neonatal history, including signs of early psychosocial difficulties. Developmental, health, and school history was asked, as well as the child's relations with his parents, siblings, and peers. Information was also obtained about the parents' backgrounds and their relations with each other and in the community. An attempt was made to learn of the child's behavior problems or any placements outside the family.

Although six hundred factors were originally listed, because of the inadequate information in some of the histories and the ambiguity of some of the items, only 83 factors were included in the final analysis. The phi coefficient, a measure of the degree of association between two dichotomous variables, was used for the statistical analysis. The lowest level of significance accepted was at the 5 per cent level.

RESULTS

Of the 83 items that were statistically analyzed, 47 were significant at the 1 per cent level and five at the 5 per cent level. These findings, as well as the statistically insignificant ones, will be discussed as far as possible in the chronological sequence of the children's development.

Prenatal and Birth History

The mothers' health during pregnancy was statistically different in the two groups, with 98 per cent of the control mothers reporting good health with no unusual complications during their pregnancies. This compared with 82 per cent of the patient mothers. Poor health ranged from prolonged nausea and vomiting to accidents, injuries, and attempted abortions. The medical history of the mothers during pregnancy was not verified by medical records or reports; however, according to the mothers' perceptions of their pregnancies, significantly more patient mothers felt that their pregnancies were characterized by ill health and unpleasantness.

Other researchers and clinicians have felt that the health and attitude of the mother during pregnancy are closely related and have a direct bearing on the early mother-child relationship. Despres found that "those (mothers) who were not happy with the fact of their pregnancy were significantly more apt to experience nausea and vomiting (and to experience nausea over a longer period of time) than women more favorably disposed toward their pregnancy" (3, p. 617). Josselyn wrote, "the emotional experiences of the mother during pregnancy have considerable bearing upon the attitude of the mother toward the newborn child, and thus upon the early relationship between the child and the mother. If she has had a happy pregnancy the baby enters a world that is, to a large measure, prepared to welcome it happily" (8, p. 28). This study did not attempt to uncover the reasons for the differences in the mothers' reports of their pregnancies. The majority of mothers in both groups, however, reported satisfactory pregnancies. From the available data, differences were not found to indicate that one group had any more abnormal or physically traumatic deliveries, although the information in this area was inadequate.

Twelve of the 71 patients (17 per cent) were born illegitimately, while none of the controls were reported born out of wedlock. Since the control group was an unusually successful sample, the patient illegitimate birth rate can be compared with a more random sample. The national illegitimate live birth rate was 4.2 per cent of all registered births in 1945 (14). The children in this study were either born in 1945 or within two years of this date. As can be seen from these figures, illegitimacy in the patient population is significantly higher than in the control group and in the national population.

The age of the parents at the time of the child's birth was not found to be statistically different in the two groups. The mean age for patient mothers was 25.3, control mothers, 26.1, and for both patient and control fathers, 29.4.

Developmental History

The age when the child began to talk in sentences was the only factor in this area that was significantly different from the available data. The controls talked in sentences an average of 1.8 months earlier (24.0 months to 25.8 months).

Differences in the average ages at time of walking, talking, and toilet training were insignificant between the two groups and within the norms suggested by Gesell and his associates (4).

Breast feeding has been thought by some to be preferable to bottle feeding for the healthiest psychosocial development of the child (9). This study did not find any significant differences in this area. About half of each group were breast fed; 48 per cent of the patients and 43 per cent of the controls were bottle fed from birth. The incidence of feeding problems in both groups was low and could not be related to either breast or bottle feeding.

Early Psychosocial Difficulties

Of interest to those concerned with the early detection and treatment of emotional and social maladjustments are the findings of early psychosocial problems in the children, as shown in Table 1.

TABLE 1

INCIDENCE OF EARLY PSYCHOSOCIAL PROBLEMS

Type of Problem	PER CENT *Patients*	 *Controls*
Enuresis*	44	14
Severe temper tantrums	35	13
Fears or phobias	25	..
Speech difficulties	23	4
Rocking or head banging	16	1
Encopresis	8	..

* Beyond 3 years of age.

With one exception—a boy who was enuretic until age 12—the controls' enuresis had stopped by school age. Parents are sometimes told not to be too concerned about problems such as enuresis, exaggerated fears and phobias, etc., as the child will "out grow" them. This is true in some cases as evidenced by the early presence of such problems among the control group; however, these findings give evidence that there is a very significant relation between these early difficulties and the later development of serious psychosocial problems.

TABLE 2

NUMBER OF EARLY PSYCHOSOCIAL PROBLEMS

Number of Problems	NUMBER Patients	Controls
1	21	18
2	16	..
3	8	1
4	..	..
5	2	..

Although some of the controls had early signs of psychosocial problems, only one had more than one such symptom. This is in contrast to the patients, 36 per cent of whom had a combination of at least two problems, as can be seen in Table 2.

The only significant cross-correlation of these early symptoms was between enuresis and severe temper tantrums. It has been suggested by several people that enuresis beyond the age of 3 years is an expression of hostility or aggression on the part of the child. These findings indicate that, if this is true, the enuretic child is likely to demonstrate his hostility and aggression by other means, particularly temper tantrums, as well. According to Allen in a study of aggression, "the boy with long standing enuresis who said, 'I want to be dry, but I don't want to be made dry,' stated the common dilemma that emerges in the struggle between the parent requiring bladder control and the child fighting against it. Enuretic behavior offers such a safe way to be aggressively oriented against responsibility and separation without the more open danger of saying, 'I will not,' to a power the child feels to be superior" (1, p. 360). If prolonged enuresis is an expression of hostility, the children in this study used temper tantrums as well as enuretic behavior to say "I will not."

The significantly high percentage of the patient group who had early rocking or head banging corroborates the theory that such behavior in the child is an early indication of the possibility of serious emotional problems in the child. Concerning rocking, Spitz has said, "Inconsistent, contradictory behavior of the mother makes the establishment of adequate object relations impossible and arrests the child at the level of primary narcissistic discharge of its libidinal drive in the form of rocking" (10, p. 111). This study did not attempt to determine the causes of rocking or head banging; therefore, the causes as stated by Spitz can neither be confirmed nor refuted.

Almost one-fourth of the patient group had early speech difficulties, such as stammering and stuttering. Another study found "a definite correlation between the development of talking and the adequacy of early care; not only did the well-mothered babies vocalize sooner but their speech maintained a smoother and easier progress" (9, p. 20). The present study did not reveal a statistically significant difference in the age at which the two groups began to talk, although the control group did begin to talk at a slightly earlier age. However, the large number of patients with speech problems was highly significant. From this study, a cause-and-effect relation cannot be established between severe emotional problems and early speech difficulties but a concomitant association is obvious.

School Problems

Because of the selection of the control subjects, they obviously did not have any serious school problems; however, school was the area of poorest performance for the patients. Academic problems in the first two grades were reported for 66 per cent of the patients; in the later grades this increased to 82 per cent. Again, this may be an important area for the early detection and treatment of problems that later develop into serious psychosocial maladjustments. This is particularly true if early school problems are accompanied by other difficulties. Academic difficulty in the first two grades was highly correlated with enuresis, speech difficulties, rocking or head banging, temper tantrums, overt aggression, and exaggerated fears or phobias.

To illustrate further the early school problems of the patients, 10 per cent were refused further admission to their schools during the first two grades, and an additional 31 per cent were refused in the later grades. Social promotions were given to 18 per cent of the patients. Truancy was a problem in 41 per cent of the patients; the difference is significant at the 5 per cent level in the first two grades and the 1 per cent level in the later grades.

Peer Relations

The ability to make friends and to get along with one's peers is considered to be healthy, normal, and desirable by most people in our society. According to their parents, the controls had no difficulties in this area; all had at least a few good friends and had no serious problems in getting along with others. The histories of the patient group showed that only 34 per cent of the patients had desirable peer relations. The desirability of the relations with their peers was decided by clinical evaluation from information obtained from parents, school, and other agency reports.

Part of their socialization pattern which could be more easily tabulated was group membership. Again, the controls made a far more satisfactory social adjustment than the patients. Each control belonged to an average of 1.7 groups while the patients' average was 0.3. Membership in one or more groups was held by almost three-fourths of the controls and only one-fifth of the patients. In other words, 80 per cent of the patients, in contrast to 27 per cent of the controls, were not members of any group. As another possible indication of the socialization patterns, 36 per cent of the patients and 86 per cent of the controls attended church at least once a month.

A study of a rural teenage population in the State of Washington found that 80 per cent of the subjects belonged to at least one social group (11). This is slightly more than the control group (73 per cent) and considerably more than the patient group (20 per cent) in this study. All of the controls lived in an urban area; however, the patients were both rural and urban. From these figures, it appears that the higher percentage of urban children in the control group does not in itself account for the differences in group memberships in the two groups.

Delinquent Behavior

Because of the bias in the selection of the control group, one would not expect to find a history of blatant delinquent behavior in the controls. Only one of the control children was reported by his parents to have been involved in delinquent activities; he was involved in a minor theft and spent one night in the juvenile detention home before being released. Forty-six (65 per cent) of the patients had been involved in delinquencies. It is interesting that more than half (27 of 46) of the delinquents committed their delinquent acts with other children rather than alone.

Of these patients with delinquency histories, 52 per cent were involved in runaways from home; 44 per cent, thefts (other than car thefts); 42 per cent, overt aggressive and assaultive behavior; 27 per cent, sexual deviancy; 22 per cent, auto theft; 18 per cent, vandalism; and 13 per cent, firesetting.

Placement Outside the Home

An overwhelming majority of the patients (72 per cent) had been placed outside their own homes prior to hospitalization. Only three of the controls had lived with anyone other than their immediate families, all three having lived with other relatives.

The patients had an average of 3.3 placements before coming to the hospital: 55 per cent had previous placements in institutions, 27 per cent with relatives, and 24 per cent in foster homes; therefore, a large number of the patients had been tried in other institutions before their hospitalization. The average length of their placements was 3.8 years. Thirty-two of these placements lasted only one year or less, while at the other extreme, two lasted for more than 10 years.

The average age of the patients at the time of first placement was 8.2 years, ranging from birth to 14 years of age. Approximately half of the first placements were caused by delinquent and incorrigible behavior on the part of the children and the other half by family break-up and parental neglect.

Medical History

The items in the medical and health histories of the children were excluded from the final analysis because of inadequate or ambiguous information. However, a history of known and reported head injuries was approximately the same in the two groups. Minor head injuries without loss of consciousness and/or hospitalization were reported in the histories of 15 per cent of the controls and 20 per cent of the patients. Head injuries with loss of consciousness and/or hospitalization were found in the histories of 15 per cent of controls and 18 per cent of the patients. Statistically insignificant was the incidence of epilepsy in the two groups.

Family Composition

Much has been written about the impact of a broken home on children's emotional and social well-being. Only 35 per cent of the patients belonged to intact families in which the child was living with both natural parents, in sharp contrast to 86 per cent of the controls. In other words, 65 per cent of the patients and only 14 per cent of the controls came from broken homes. Thirteen patients and four controls had stepparents, and the remaining five patients and three controls were living with their adoptive parents.

In addition, the marriages of the controls' natural parents had lasted longer than had the marriages of the patients' natural parents. The mean duration of the natural parents' marriage was 18.8 years for the controls and 12.0 for the patients.

The age of the child at the time of the break-up of his natural family is of interest. The patients had a mean age of 4.4 years when they were first separated from one or both of their natural parents, approximately the same as in the control group. Of the nine control children from broken families, five were over 4 years of age at the time of the break-up; the other four were 4 years of age or younger.

The number of patients from broken homes was considerably higher than those found in a study of less seriously disturbed subjects. Ingham, in his study of out-patient neurotics, found that 31.1 per cent of the neurotic patients came from homes broken by separation and 1.8 per cent from homes broken by death. The total 32.9 per cent had their homes broken prior to their twenty-second birthday, an age considerably older than the mean age of 4.4 for the patients in this study. A study done in the State of Washington in nine rural public high schools (13) found that 80.7 per cent of the students came from intact homes, a figure just slightly lower than this study's urban control group.

The size of the family in the patient group was significantly larger. The families of the patients had an average of 4.0 children compared to 2.8 in the families of the controls. Hawkes, Burchinal, and Gardner concluded ". . . that a small family environment does not have a detrimental effect on the children's personality development. . . . there was a suggestion that children from smaller families might fare better psychologically than children from larger families. Certainly studies employing a larger sample or ones in which important variables such as socio-economic status and perhaps age differences among the children have been controlled need to be initiated" (6, p. 68). The sample in this study was not larger than the Iowa study but age was controlled and socioeconomic status was approximated. The control children, the group that came from smaller families, obviously "fared better psychologically" since the patient group was seriously enough disturbed to warrant psychiatric hospitalization.

The ordinal position of the children in the two groups was not significant. Six in each group were the only child in their families. An interesting but not significant comparison of the two groups was that more patients than controls were middle children, ranging from next to the oldest to next to the youngest. Thirty-five controls and 25 patients were the oldest children in their families, and 17 controls and 11 patients were the youngest.

Mothers' History

Several significant items differentiated the patients' and controls' mothers. The following problem areas were found in the histories of the patients' mothers: hospitalization for mental illness, 7 per cent; epilepsy in the immediate family, 10 per cent; excessive drinking, 15 per cent; and chronic physical illness with or without hospitalization, 10 per cent. None of the controls' mothers reported these problems.

The control mothers had a higher level of education, 11.5 to 10.7 years. In the control group 48 mothers had completed the twelfth grade and 21 of these had some college or vocational training. On the other hand, only 32 of the patients' mothers had completed the twelfth grade and only seven had additional education or training.

There was no statistically significant difference between the ages of the mothers at the time of their marriage or at the time of the birth of the child studied. The average age at marriage was 21 years for both groups, and the age at the time of the child's birth has been given in a previous section of the paper. The average number of marriages did not prove to be statistically significant, 1.1 for the control mothers and 1.7 for the patient mothers.

Working mothers have been the topic of many discussions in both professional and popular literature. Many mothers who want or have to work outside the home have felt guilty because of doing so. In this study only slightly more of the patients' mothers worked outside the home, 33 to 25. However, the patients' mothers earned significantly more income per month, $202 compared with $181. Since the controls' mothers had more education and training, speculation might attribute the higher income of the patients' mothers to longer working hours. Of interest in these findings is a conclusion of the Glueck and Glueck report: "Equal proportions of mothers of nondelinquents and delinquents were regularly employed but a greater proportion of mothers of delinquents than nondelinquents worked irregularly" (5, p. 330). From the available data in this study, the regularity of the mothers' employment could not be determined.

The history of mental deficiency in the mothers and their families failed to show any statistical significance as did the mothers' criminal or court records.

Fathers' History

One of the significant differences in the two groups of fathers was their educational attainment. As with the mothers, the controls' fathers averaged more education, 11.5 to 10.5 years. In the control group 45 fathers had completed high school and 15 had further education or training. In the patient group the figures were 23 and 10 respectively. The controls' fathers reported no history of problems with alcohol or arrests. In contrast, the histories of the patients' fathers revealed that 14 per cent had a criminal or court record and 18 per cent had a history of excessive drinking.

The mean occupational level of the two groups of fathers was not statistically different. Most of the fathers in both groups worked as operatives or craftsmen. However, certain differences in the job classifications were found. Grouping the fathers by the U. S. Census Bureau occupational classifications, the frequency of the patient fathers in the extreme categories, professional-managerial and service-labor, was significant. The loading of the patients' fathers in the extreme categories may be due to the classification system used. For instance, included in the professional group of the patients' fathers was a jockey with an unstable work history.

The control fathers had a significantly higher income per month, $464 to $348. A more stable work history for the control fathers was revealed in the findings and may account for the higher income in spite of the similarities of the mean occupational levels of the two groups.

Approximately the same number of control fathers (44 per cent) and patient fathers (37 per cent) had been in the armed forces. The available data did not provide more detailed information about their service careers. In addition, the two groups of fathers were not significantly differentiated by the following items: hospitalization for mental illness, epilepsy in their immediate families, a history of mental deficiency, number of marriages, and age at time of marriage. The average number of marriages was 1.3 for the patients' fathers and 1.1 for the controls' fathers. There was only a one year difference in their average ages at time of marriage: patient fathers, 25.6; control fathers, 24.6.

Family Socioeconomic Status

The data did not lend itself to the establishment of socioeconomic status by the use of such criteria as established by Warner, Meeker, and Eells (12). However, certain conclusions can be drawn from the available data. Not only did the control fathers earn more than the patient fathers, but also the control families' total family income was significantly higher. The mean family income for the control group was $513 per month and exceeded the patient group by exactly $100. The mean family income for both groups was higher than that for the nation in 1955, which was $387 per month (14). Although the patients' mothers earned more than the controls' mothers, the difference in the fathers' income was sufficient to give the controls a higher family income. Although both groups had a family income higher than the national average, the control families had a higher income with fewer children to support.

It has already been mentioned that the mean occupational level of the fathers was similar. Although Warner found that educational level has a very low correlation with other socioeconomic indices, other studies have used educational attainment as a criterion for establishing socioeconomic status (2). As had already been reported, the educational level of both parents was significantly higher in the control group of this study.

At the time the histories were taken, 20 per cent of the patient families and none of the control families were receiving financial assistance from a public agency. The assistance received by the patients' families was usually from the Aid to Dependent Children program. An additional 10 per cent of the patient group and 1.4 per cent of the control group had received public assistance in the past but were self-supporting at the time of the study.

Since many of the patients came from families with multiple problems, it was not surprising that 30 per cent had a history of financial problems severe enough to warrant public assistance. Perhaps more surprising is that one-third of the patient families who had ever received public assistance were self-supporting at the time of the study. Equally surprising may be the fact that one of the successful control children belonged to a family that had received public assistance in the past.

Integration of Families in the Community

Participation in community organizations clearly distinguishes the patient and control families. Of the control families, 87 per cent were actively associated with at least one community organization in contrast to only 22 per cent of the patient families. Church attendance at least once a month was reported by 58 per cent of the control families and 38 per cent of the patient families. Significantly more controls than patients attended church with their parents.

Home ownership also differentiated the two groups: 89 per cent of the control families compared to 47 per cent of the patient families owned or were buying their homes.

DISCUSSION

The limitations of this study should be examined before evaluating the findings. Parents of both groups knew why their children were being studied, and this undoubtedly had an influence on the information that they gave the interviewers. The parents of the patients were well aware that their children were in a state mental hospital and considered unsuccessful by many people. Anyone who has interviewed the parents of emotionally disturbed children should be aware of the feelings, such as guilt, denial, rationalization, and projection, that the parents bring with them to the interview, as well as the fear of being blamed or criticized. The possible coloring and distortion of factual information by such feelings is well-known to all in the mental health fields.

Although the information given by the patients' parents was often compared with reports from other agencies, this was not possible in all cases. In addition, other agency personnel, sometimes oriented to look for pathology, knew that the child had been admitted to a state mental hospital when their reports were requested. Therefore, it is possible that agency reports emphasized the pathological aspects of the child's history and behavior.

On the other hand, the parents of the control children knew that their children were chosen for study because they were "successful." It is impossible to know the effect of this knowledge on their perceptions and descriptions of the children's histories. Both control parents were not always seen for the history interview; therefore, the recollections and the opinions of just one parent often had to be accepted as valid.

Another factor that affected the study's findings was the inadequacy of information in the patients' records. Not all of the patients' parents were interviewed. Often just one parent gave the social history, usually the mother. The patients' parents were less able to remember and relate parts of the history than were the controls' parents. Since many of the patients had been placed outside their parents' homes, the parents lacked firsthand information about certain items in their histories. In some cases, agency reports were able to fill these gaps.

The interviewers knew that the child was in either the patient or control group before they took the history. This information may have biased the interviewer in the type of information he obtained and the interpretation of the findings. Although the histories of the control subjects were taken by professionally trained psychiatric social workers, the histories of the patients sometimes were taken by less trained personnel.

Therefore, because of a possible bias in the perceptions of both the interviewers and the parents, pathology may have been emphasized in the histories of the patients and normalcy may have been emphasized in the histories of the controls.

Still other chances for human error should be considered. Several different people were involved in the collection and tabulation of the data. While this has its advantage in eliminating the personal bias of just one person, it has its disadvantages in lack of uniformity and continuity.

Although many statistically significant differences were found between the patient and control groups, few of the items were found to occur exclusively in either group. The two groups were selected because of their extreme difference in social adjustment; therefore, significant differences in their social histories would necessarily be expected. The causes of the differences and similarities in the two groups were not investigated in the study. More detailed information would be desirable for several of the items but was beyond the scope of the study.

In spite of its limitations, the study does contribute some reliable findings of interest and significance. It almost goes without saying that many of the areas discussed in this paper need and warrant further investigation. Perhaps one of the greatest values of statistical studies is their ability to indicate areas for more intensive research.

CONCLUSIONS

Similarities in the Two Groups

A comparison of the social histories of the patient and control groups reveals some interesting similarities. The parents of both groups were about the same age not only at the time of their marriage but also at the birth of the child studied. A majority of the parents in either group had been married only once. The percentage of working mothers in both groups was approximately the same, and slightly less than half of all the fathers had been in the armed forces.

In addition, approximately half of the children in each group were breast fed and half were bottle fed from birth. Feeding problems were rare in both groups. The average ages at time of walking, talking, and toilet training of all the children were within normal limits and about the same in both groups. Very few of the children were the only child in the family, and their ordinal positions were similar. A very small percentage of the children in each group was adopted. Although the health histories of the children were inadequate, head injuries were reported to about the same degree in the two groups.

The Patient Group

A composite picture of the patient group shows some interesting contrasts to the control group. Although a majority of all the mothers reported good health during their pregnancies, more of the patients' mothers told of having difficult or complicated pregnancies. Over 10 per cent of the patients' births were illegitimate. The ability to talk in sentences came more slowly for the patients. Early psychosocial problems, such as enuresis, severe temper tantrums, fears and phobias, speech difficulties, rocking, and head banging, were prominent during the patients' infancy and early childhood. The children who were enuretic also tended to have severe temper tantrums. Those children who had early psychosocial problems also had difficulty in their school performance. School achievement and adjustment were serious problems for the patient group and appeared in the early years of their school histories.

A large majority of the patients had poor relations with their peers and did not belong to any social group outside their family and classroom. More than half of the patients were involved in delinquent behavior, and they tended to get involved in delinquencies with other children rather than alone. Almost three-fourths of the patients had been placed in foster homes or other institutions prior to their hospitalization with an average of over three such placements.

An overwhelming majority of the patients came from broken homes. The patients also came from larger families than did the controls. Patients' families had lower family incomes, and their fathers had a more unstable work history. Although only slightly more patients' mothers worked outside the home, they earned significantly more than the controls' mothers. Public assistance was much more often needed by the patients' families. One-third of those receiving such help eventually became self-supporting.

Both parents of the patients had an average of less than an eleventh grade education. The mothers of the patients were more likely to have been in mental hospitals and to have had problems with alcohol. One-third of the patients' mothers had a history of epilepsy in the immediate families. The patients' fathers were differentiated by court and criminal records, as well as excessive drinking. Patient families were not well integrated in the community and took little part in community organizations. They also tended to live in rented family dwellings.

The Control Group

The successful control child was born of married parents, and his mother's pregnancy was healthy and without complications. This child talked in sentences at an earlier age. Early signs of psychosocial problems were an exception rather than a rule. If the child had been enuretic at all, the enuresis stopped before he started to school. The control children had no serious problems in school or in their ability to get along with other children. Participation in social groups outside the family and classroom was common for these children, as was church attendance. Delinquent behavior was virtually unknown among this group. The successful child had never experienced the trauma of living with anyone other than his relatives and was very likely to be living with both natural parents.

The control family was smaller and had a significantly higher family income. A more stable work history and a higher income were found among the control fathers. The need for public assistance was very rare among this group of families. Although the parents of these children averaged less than a high school education, their educational level was higher than the parents of the patients. Family histories of mental illness and problems with alcohol were not reported by the control parents. These parents were active in community and religious organizations and were likely to own their homes or be in the process of buying them.

SUMMARY

The social histories of children in a state mental hospital were compared with those of successful children in a control group, matched for sex and age. A total of 142 children, between the ages of 13 and 16 years, were included in the study. Of the 83 items in the histories that were statistically analyzed, 52 were significant at or beyond the 5 per cent level. The differences and similarities in the social histories of the patient and control groups were presented and discussed. Some traditionally held theories were confirmed and others were refuted or questioned by the findings of this study.

REFERENCES

1. ALLEN, F. Aggression in relation to emotional development, normal and pathological. *Ment. Hyg.*, 1950, 34, 353-363.
2. BURCHINAL, L., GARDNER, B., & HAWKES, G. R. Children's personality adjustment and the socioeconomic status of their families. *J. genet. Psychol.*, 1958, 92, 149-159.
3. DESPRES, M. Favorable and unfavorable attitudes toward pregnancy in primiparae. Unpublished Ph.D. dissertation, Univer. of Chicago, 1936. (Secondary source: WALLIN, P. & RILEY, R. P. Reactions of mothers to pregnancy and adjustment of offspring in infancy. *Amer. J. Psychiat.*, 1950, 20, 616-622.)
4. GESELL, A., & ILG, F. L. *The child from five to ten.* Harper, 1946.
5. GLUECK, S, & GLUECK, E. Working mothers and delinquency. *Ment. Hyg.*, 1957, 41, 327-352.
6. HAWKES, G. R., BURCHINAL, L., & GARDNER, B. Size of family and adjustment of children. *Marriage Fam. Living,* 1958, 20, 65-68.
7. INGHAM, H. V. A statistical study of family relationships in psychoneurosis. *Amer. J. Psychiat.*, 1949, 106, 91-98.
8. JOSSELYN, I. *Psychosocial development of children.* Family Service Association of America, 1948.
9. RIBBLE, M. *The rights of infants.* Columbia Univer. Press, 1943.
10. SPITZ, R. Autoeroticism: some empirical findings and hypotheses on three of its manifestations in the first year of life. *Psychoanal. Stud. Child*, 1949, 3-4, 85-119.
11. STONE, C. L. A summary of the responses of Ferry County teenagers to a questionnaire on activities and social relationships. Unpublished ms., Washington State Univer., 1958.
12. WARNER, W. L., MEEKER, M., & EELLS, K. *Social class in America.* Science Research Associates, 1949.
13. Responses of Stevens County high school students to a questionnaire on activities and social relationships of teenagers. Unpublished ms., Dept. of Rural Sociology, Washington State Univer., 1958.
14. Statistical Abstract of the United States, 1957. Bureau of the Census, U. S. Dept. of Commerce, U. S. Gov't Printing Office, Washington, D.C., 1957.

CHILD BEHAVIOR PROBLEMS AND PARENTAL ATTITUDES[1]

DONALD R. PETERSON,* WESLEY C. BECKER, DONALD J. SHOEMAKER, ZELLA LURIA,[2] *and* LEO A. HELLMER
University of Illinois

If modern personality theory and clinical practice have any basis in fact, the importance of parental attitudes in the origin of behavior problems among children can be assumed. The nature of the relationships between parent and child behavior, however, cannot merely be assumed, but must either be derived from or confirmed by empirical research. For some time, the investigators have been engaged in a program of study designed to clarify certain associations between parental characteristics and behavior tendencies of children. The work parallels a great deal of previous research in this area, but differs from most of the latter in the systematic inclusion of variables derived from examination of fathers. In most contemporary research, and in the usual clinical management of child behavior problems, emphasis on the maternal influence is overwhelmingly strong. Consequent neglect of the attitudes and other qualities of fathers has, until now, been justified only by assumption or convenience, and the assumption has seemed to us to require test.

Two studies (1, 5), involving the examination of characteristics of both parents in reference to certain behavior tendencies among children, offered such a test and yielded, among others, the following results:

1. Contrary to general assumption, the attitudes of fathers were at least as intimately related as the attitudes of mothers to the occurrence and form of maladjustive tendencies among children.

---

* Department of Psychology, University of Illinois, Urbana.

[1] This study was supported by the Department of Public Welfare, State of Illinois. We extend our gratitude to the participating parents and teachers, especially of Westview and Bottenfield Kindergartens in Champaign, for their part in providing the data, our high praise to Herbert Quay and Rodgers Elliott for their management of the many administrative problems of the study, and our appreciation to Irvin Moelis and Geraldine Piorkowski for their help in the data analysis.

[2] Now at Tufts University.

Child Develpm., 1961, 32, 151-162.

2. Personality problems (shyness, withdrawal, etc.) among children attending a guidance clinic were relatively independent of maternal attitudes, but were related to autocratic attitudes and lack of parental concern among fathers.

3. Conduct problems (truancy, stealing, etc.) were associated with general maladjustment among mothers and with evident permissiveness and disciplinary ineffectuality on the part of fathers.

Like any such findings, these permit certain conclusions to be drawn about parent-child relationships, but generalizations are limited by the specificity of the methods used and the particular samples chosen for investigation. For this reason it is heuristically advantageous to regard such results as guiding hypotheses for further research, and this is what we have done in the present investigation. With improved, but comparable, methods of measuring parental attitudes and child behavior and with differently selected samples of subjects, we have sought to assess the generality of the above propositions and thus to evolve more substantial conclusions about the way parents and children affect each other.

SUBJECTS

The study deals with two groups of families. Children in the first group were drawn from two local kindergartens solely on the basis of willingness to participate in the research. We shall call them the *Kindergarten* sample. Families in the second group had come to the University of Illinois Psycho-

TABLE 1

SAMPLE CHARACTERISTICS

	KINDERGARTEN GROUP		CLINIC GROUP	
	Mean	*SD*	*Mean*	*SD*
Age of child (in months)	65.3	2.9	69.8	7.1
Age of father (in years)	35.3	6.0	33.7	4.6
Age of mother (in years)	33.0	5.2	31.5	3.5
Education of father (grades completed)	16.0	2.6	15.1	2.7
Education of mother (grades completed)	14.3	2.0	13.0	1.5
Sex of child:				
Number of boys	24		17	
Number of girls	29		7	
Occupation of father (Warner):				
Level 1 (professional, etc.)	9		5	
Level 2	30		4	
Level 3	10		8	
Level 4	2		2	
Level 5 (skilled labor, etc.)	2		5	

logical Clinic to deal with one kind of child behavior problem or another. Let us call them the *Clinic* sample. An important consideration in forming the latter group was to select children of approximately kindergarten age, but, beyond that, few restrictions were imposed. Cases of obvious intellectual defect and of known organic disability were not included, nor were those in which the parents were separated or unavailable for study. All subjects were white. An effort was made to exclude families involving adoptive relationships of any sort, but one such family was erroneously admitted anyway and retained for subsequent analyses. By the time the mistake was discovered, it was too late to do anything about it. Various characteristics of the samples are described in Table 1.

The Kindergarten sample, particularly, and both samples, to some extent, represent a restricted upper range of the national population in respect to socioeconomic, educational, and intellectual status.[3] The generality of the study is therefore limited, but, to make a virtue of necessity, the reduction in variance along such extraneous dimensions as intelligence should attenuate relationships between them and the factors with which the study is centrally concerned.

In a clinic the size of ours, the number of kindergarten age children admitted in any given year is small, and we were forced to take practically all such clients who came along. As a result, the quality of control leaves something to be desired. The Kindergarten and Clinic groups differ in respect to sex of child (there are over twice as many boys as girls in the Clinic group), age of child (Clinic children are older, on the average, and vary more in age than those in the Kindergarten group), education of mother (Kindergarten mothers went to school longer), and occupation of father (occupational level was higher for fathers in the Kindergarten group). The possible effect of these differences on subsequent comparisons is difficult to assess, but an effort to do so was made by correlating each of the background variables with parental attitudes as we measured them. Only two of the 40 pertinent relations resulting from that analysis were "significant" at the .05 level. The generally low magnitude of the correlations, and the fact that many of the relations were such as to reduce rather than to elevate the effects we shall discuss below, suggests that differences in background were not crucial factors in determining results.

PROCEDURES

The structured interview developed by Sears, Maccoby, and Levin (6) was modified slightly to permit use with fathers and administered very much as Sears *et al.* have described it with mothers. All interviews were recorded, and two judges rated each subject in respect to various parental attitudes and practices. From these ratings, scores were obtained for five of the seven factors which Sears and his colleagues derived in the course of their study. Unit weights were employed throughout. The elemental variables comprising each factor and the factor designations themselves are set out below:

---

[3] No measures of intelligence were obtained, but school records showed that the mean IQ for children in the two kindergartens regularly approximates 120.

Factor 1. *Strictness.* Restrictions on play, demands for orderliness, demands in regard to table manners, severity of toilet training, standards of obedience, emphasis on scholastic achievement, (negative) response to dependency, use of physical punishment, restrictions on and punishment for aggression against parents and siblings, restrictions on masturbation, nudity, and sex play.

Factor 2. *Adjustment.* Self-esteem, esteem for spouse, satisfaction with life situation, (low) child-rearing anxiety, delight of self and spouse over pregnancy, (little) disagreement over child-rearing, report of affection between spouse and child, tendency to follow through on disciplinary threats and promises.

Factor 3. *Warmth.* Warmth, acceptance of dependency, use of praise and reasoning in discipline.

Factor 4. *Responsibility* Self-esteem, demands in respect to orderliness and table manners, restrictions on play, use of praise when children play nicely, standards of obedience, deprivation of privileges, use of tangible rewards in discipline.

Factor 5. *Aggression.* Demands that child "fight back," (low) permissiveness and high punishment for aggression against parents, use of physical punishment in discipline.

Independent evaluations of child behavior were obtained from parents and teachers, using a problem checklist developed for use in this investigation. To select a nonarbitrary sample of problems, 427 representatively chosen case folders for children attending the Clinic were inspected, and the referral problems of each child were recorded. Principally on the basis of frequency of occurrence, 58 problems were selected for general investigation. Both parents and two teachers rated each of the Kindergarten children who participated in the study, and both parents and one teacher rated the children in the Clinic group. In addition, the teachers completed checklists for all nonparticipating subjects in one of the kindergartens. The latter data ($N = 62$) were factor analyzed to permit specification of reasonably unitary, independent dimensions of problem behavior. Since three of the original 58 problems (soiling, enuresis, vertigo) were never observed, they were excluded from further consideration. Phi coefficients of intercorrelation were computed for the remaining 55 variables, and 10 factors extracted by the complete centroid method. Since the first three factors accounted for 55 per cent of the common variance represented in the 10 factors examined and none of the rest of the factors accounted for more than 10 per cent of that variance, only the first three were retained for rotation. To exclude any possibility of subjective bias in defining problem dimensions, Kaiser's varimax criterion (4) was employed as the sole basis for determining vector placement. Factor loadings are presented in Table 2. The first two factors seem rather clearly to represent the conduct problem and personality problem dimensions of concern in our previous work and have been named

TABLE 2

ROTATED FACTOR MATRIX

	FACTOR LOADING		
	1	2	*3*
1. *Conduct Problem*			
Disruptiveness	70	—17	15
Boisterousness	70	—30	08
Disobedience	68	—01	02
Attention-seeking	66	—26	—05
Uncooperativeness	62	—05	15
Restlessness	60	—18	28
Fighting	59	—09	06
Negativism	59	15	—12
Impertinence	59	04	—23
Irritability	55	04	—16
Destructiveness	52	—12	28
Distractibility	52	16	38
Irresponsibility	51	05	39
Hyperactivity	50	—15	—06
Temper tantrums	49	28	—07
Laziness	41	08	28
Tension	39	28	—11
Profanity	32	—15	07
Dislike for school	30	23	—28
Thumb-sucking	29	—04	15
Headaches	26	26	—17
Jealousy	23	10	—19
Specific fears	—19	00	16
2. *Personality Problem*			
Inferiority feelings	—02	58	27
Tendency to be flustered	—07	58	12
Lack of self-confidence	—09	57	26
Anxiety	06	55	—07
Self-consciousness	—28	52	05
Aloofness	—19	52	—09
Shyness	—32	51	21
Reticence	10	49	14
Social withdrawal	—20	47	08
Inability to have fun	—11	45	02
Hypersensitivity	16	42	—37
Depression	—03	39	07
Lack of interest in environment	05	39	16
Nervousness	35	36	—16
Crying	28	35	—29

(*table continued on next page*)

TABLE 2 (*continued*)

ROTATED FACTOR MATRIX

	Factor Loading *1*	Factor Loading 2	Factor Loading *3*
Preference for older playmates	02	—28	—03
Skin allergy	—03	—22	—01
3. *Autism*			
Preoccupation	—04	20	50
Shortness of attention span	29	22	50
Clumsiness	08	27	50
Inattentiveness	39	24	47
Passivity	01	27	46
Daydreaming	03	38	44
Sluggishness	—20	23	40
Stuttering	06	06	39
Hay fever or asthma	08	—02	37
Drowsiness	—07	20	36
Preference for younger playmates	15	—01	36
Abdominal pain	12	16	—25
Nausea	03	02	—25
Masturbation	10	11	—17
Truancy from school	10	11	—17

accordingly. The third factor is more difficult to interpret, but seems to contain a preponderance of autistic elements and has been labeled to conform with that inference. Its meaning is not unlike that of the tentatively identified "prepsychotic" factor isolated by Himmelweit (3) and described by Eysenck (2). Factor scores for the subjects who participated in the research were computed by assigning unit weights to the 10 most heavily saturated variables in each of the three dimensions.

Analysis of relations between parental attitudes and child behavior proceeded in two ways. First, the attitudes of Kindergarten and Clinic parents were compared. Distributions of the five attitude scores were formed separately for the mothers and fathers in the Kindergarten group, and cutting points established as close as possible to the first and third quartiles in each of the distributions. Attitude scores for Clinic parents were then tabulated on the same baseline, and χ^2 tests of significance performed. In the second analysis, correlations between parental attitudes and child problem factors were determined for the Kindergarten and Clinic groups. Three of the 24 Clinic children were not in school, teacher ratings were therefore unobtainable, and the pertinent correlations reported in Table 5 below are based on an *N* of 21 cases.

TABLE 3

INTERJUDGE CORRELATIONS

Parent Attitude	*Mothers*	*Fathers*	*Child Problem*	*Inter-Parent*	*Inter-Teacher*	*Parent-Teacher*
Strictness	.74	.61	Conduct problem .	.48	.82	.41
Adjustment ...	.70	.72	Personality problem	.52	.68	.24
Warmth	.57	.68	Autism	.62	.74	.41
Responsibility .	.59	.45				
Aggression ...	.81	.81				

Interjudge correlations for the various factors are given in Table 3. These were computed for the Kindergarten sample only, since only in that sample were data from two raters available for all measures. With the exception of the teacher ratings for Clinic subjects, all variables used in the correlational analyses represented average ratings of two judges. The *r*s as reported can therefore be corrected by means of the Spearman-Brown formula. When this is done, interjudge consistencies range from .64 for parent ratings of conduct problem to .90 for teacher ratings of the same dimension. The mean corrected *r* over all reliabilities pertinent to the analysis is .80.

Correlations between parent ratings and teacher ratings of child behavior, however, are relatively low. It would have been convenient and parsimonious to construe these problem dimensions as general, trans-situational tendencies, but to do so seemed operationally unwarranted. Ratings by parents and by teachers were therefore considered separately in all analyses.

RESULTS

Results of the comparison of Kindergarten and Clinic parents are given in Table 4. An inordinate share of both mothers and fathers in the Clinic group were described as poorly adjusted and lacking in warmth. In addition, a significant proportion of the Clinic mothers were said to be irresponsible, and a relatively large number of the Clinic fathers were described as strict and aggressive. The composite attitudinal picture for Clinic mothers, therefore, is of a cold, maladjusted woman, evidently unable or unwilling to assume responsibility for the rearing of her children. The modal Clinic father appears to be a cold, maladjusted, punitive man, lacking the human kindness shown by the fathers of the relatively well-adjusted kindergarten children.

Relations between judgments of parent attitudes and child behavior are presented in Table 5. The most striking findings for the Kindergarten

TABLE 2 (*continued*)

ROTATED FACTOR MATRIX

	FACTOR LOADING		
	1	2	*3*
Preference for older playmates	02	—28	—03
Skin allergy	—03	—22	—01
3. *Autism*			
Preoccupation	—04	20	50
Shortness of attention span	29	22	50
Clumsiness	08	27	50
Inattentiveness	39	24	47
Passivity	01	27	46
Daydreaming	03	38	44
Sluggishness	—20	23	40
Stuttering	06	06	39
Hay fever or asthma	08	—02	37
Drowsiness	—07	20	36
Preference for younger playmates	15	—01	36
Abdominal pain	12	16	—25
Nausea	03	02	—25
Masturbation	10	11	—17
Truancy from school	10	11	—17

accordingly. The third factor is more difficult to interpret, but seems to contain a preponderance of autistic elements and has been labeled to conform with that inference. Its meaning is not unlike that of the tentatively identified "prepsychotic" factor isolated by Himmelweit (3) and described by Eysenck (2). Factor scores for the subjects who participated in the research were computed by assigning unit weights to the 10 most heavily saturated variables in each of the three dimensions.

Analysis of relations between parental attitudes and child behavior proceeded in two ways. First, the attitudes of Kindergarten and Clinic parents were compared. Distributions of the five attitude scores were formed separately for the mothers and fathers in the Kindergarten group, and cutting points established as close as possible to the first and third quartiles in each of the distributions. Attitude scores for Clinic parents were then tabulated on the same baseline, and χ^2 tests of significance performed. In the second analysis, correlations between parental attitudes and child problem factors were determined for the Kindergarten and Clinic groups. Three of the 24 Clinic children were not in school, teacher ratings were therefore unobtainable, and the pertinent correlations reported in Table 5 below are based on an N of 21 cases.

TABLE 3

INTERJUDGE CORRELATIONS

Parent Attitude	*Mothers*	*Fathers*	*Child Problem*	*Inter-Parent*	*Inter-Teacher*	*Parent-Teacher*
Strictness	.74	.61	Conduct problem .	.48	.82	.41
Adjustment . . .	.70	.72	Personality problem	.52	.68	.24
Warmth	.57	.68	Autism	.62	.74	.41
Responsibility .	.59	.45				
Aggression . . .	.81	.81				

Interjudge correlations for the various factors are given in Table 3. These were computed for the Kindergarten sample only, since only in that sample were data from two raters available for all measures. With the exception of the teacher ratings for Clinic subjects, all variables used in the correlational analyses represented average ratings of two judges. The *r*s as reported can therefore be corrected by means of the Spearman-Brown formula. When this is done, interjudge consistencies range from .64 for parent ratings of conduct problem to .90 for teacher ratings of the same dimension. The mean corrected *r* over all reliabilities pertinent to the analysis is .80.

Correlations between parent ratings and teacher ratings of child behavior, however, are relatively low. It would have been convenient and parsimonious to construe these problem dimensions as general, trans-situational tendencies, but to do so seemed operationally unwarranted. Ratings by parents and by teachers were therefore considered separately in all analyses.

RESULTS

Results of the comparison of Kindergarten and Clinic parents are given in Table 4. An inordinate share of both mothers and fathers in the Clinic group were described as poorly adjusted and lacking in warmth. In addition, a significant proportion of the Clinic mothers were said to be irresponsible, and a relatively large number of the Clinic fathers were described as strict and aggressive. The composite attitudinal picture for Clinic mothers, therefore, is of a cold, maladjusted woman, evidently unable or unwilling to assume responsibility for the rearing of her children. The modal Clinic father appears to be a cold, maladjusted, punitive man, lacking the human kindness shown by the fathers of the relatively well-adjusted kindergarten children.

Relations between judgments of parent attitudes and child behavior are presented in Table 5. The most striking findings for the Kindergarten

TABLE 4

COMPARISON OF KINDERGARTEN AND CLINIC PARENTS

	MOTHERS			FATHERS		
Parent Attitude	*Kindergarten*	*Clinic*	χ^2	*Kindergarten*	*Clinic*	χ^2
Strictness			.75			6.81*
high	14	6		13	13	
average ..	26	14		26	6	
low	13	4		14	5	
Adjustment ...			8.43*			23.26*
high	12	1		13	1	
average ..	30	11		27	3	
low	11	12		13	20	
Warmth			14.65*			11.67*
high	16	1		14	1	
average ..	23	6		26	8	
low	14	17		13	15	
Responsibility .			14.02*			1.56
high	14	1		15	5	
average ..	23	5		25	15	
low	16	18		13	4	
Aggression ...			.31			9.77*
high	16	7		11	13	
average ..	21	11		31	6	
low	16	6		11	5	

* Significant at or beyond the .05 level.

sample are those which indicate covariation between strictness, lack of warmth, and aggression among fathers and problems of all kinds among the children. Contrary to expectation, the various problem dimensions were not differentially associated with specific attitudinal patterns. Conduct problems, personality problems, and autistic tendencies were all related in much the same way to the cluster of punitive attitudes mentioned above. However, there appeared to be some specificity in regard to the situation in which the problem behavior was observed. Strict, cold paternal attitudes, for instance, were most closely related to personality problems as judged by parents, but were more highly correlated with conduct problems as teachers observed them in school. It is entirely possible that strict, cold fathers simply will not allow emotional display in the home. A considerable amount of tension would probably result from sustained suppression of this

TABLE 5

CORRELATIONS BETWEEN PARENTAL ATTITUDES AND CHILD PROBLEMS
(Decimal Points Omitted)

	MOTHERS						FATHERS					
	Conduct Problems		*Personality Problems*		*Autism*		*Conduct Problems*		*Personality Problems*		*Autism*	
	PR†	TR‡	PR	TR	PR	TR	PR	TR	PR	TR	PR	TR
KINDERGARTEN SAMPLE (*N* = 53)												
Strictness . . .	11	09	−13	10	20	13	16	34*	34*	08	50*	16
Adjustment .	04	−12	−18	−20	−06	−14	−17	−15	−26	04	−22	−03
Warmth	05	04	20	−23	−01	−25	−17	−40*	−30*	−15	−24	−08
Responsibility	14	13	13	−12	32*	−01	08	07	26	09	33*	07
Aggression . .	16	14	−14	19	25	27*	11	20	34*	25	40*	17
CLINIC SAMPLE (*N* = 21 − 24)												
Strictness . . .	07	23	−43*	−49*	−25	−29	38	23	−19	−33	04	−02
Adjustment .	−35	−34	−32	−15	−36	−20	−29	−18	−35	−20	−02	00
Warmth	−06	−24	14	30	23	21	−01	16	15	−04	23	24
Responsibility	25	01	−38	−54*	−09	−39	38	29	−34	−50*	15	07
Aggression . .	32	25	−19	−37	05	−17	23	57*	−10	−25	10	10

* Significant at or beyond the .05 level.
† Parent ratings of child problem.
‡ Teacher ratings of child problem.

kind, to find outlet as conduct disturbance in the freer atmosphere of kindergarten.

The situational difference discussed above is an exception rather than a rule, and, in general, correlations between parental attitudes and teacher ratings of child behavior are similar to those between parental attitudes and parent ratings of child behavior. Thus, the pattern of parental influence seems about the same whether children are observed in school or at home, even though teachers and parents view children from somewhat different vantage points. In retrospective light of these results, the low correlations between parent and teacher ratings do not seem as restrictive as they did when the analyses were begun. At the operational level, the results provide some justification for pooling such judgments when future analyses are planned, and, at the theoretical level, for conceiving once more of general child problem tendencies pervading behavior in many different kinds of situations.

Correlations for the Clinic families must be interpreted much more cautiously than those for the Kindergarten group. For a sample of 24 cases, a correlation must exceed .42 to achieve significance at the .05 level, and correlations of this magnitude are necessarily rare in the study of parent-child relationships. The data were also vulnerable to contamination from knowledge of child problem status and are presented more in the interest of completeness and integrity than because of any new truths they reveal. Most of the "significant" correlations are between parental strictness and responsibility and the *absence* of personality problems among the children, possibly representing a relaxation of pressure by parents who perceive shyness, withdrawal, and other personality problems in their offspring. No other relations were confirmed, and most conclusions drawn from the study must be justified through reference to the comparison between Clinic and Kindergarten families and to the relatively "pure" correlations established within the larger Kindergarten sample.

Discussion

Let us now restate the hypotheses which guided this research and consider the results in the light of those propositions.

Hypothesis 1. The attitudes of fathers are at least as intimately related as the attitudes of mothers to maladjustive tendencies among children.

The prediction implicit in this hypothesis was clearly borne out. Of the 65 relations examined for fathers, 14 were "significant" at the .05 level. Among the same number of relations examined for mothers, only eight were "significant." One should not make too much of that disparity because of the multiple interdependencies among the variables, but the direction of the difference, at least, is interesting. The paternal influence, furthermore, appears just as strongly predominant for the present Kindergarten group as it has for previous samples (1, 5). This is important not only because data from the Kindergarten group are almost wholly free from contamination, but because the sexes are almost equally represented in the group. In fact, a few more kindergarten girls than boys participated in the study. Reservations arising from the preponderance of male children in groups studied before are thus mollified somewhat, and the critical importance of the father role is clearer than ever.

Hypotheses 2 and 3. Child personality problems are related to autocratic attitudes and lack of parental concern among fathers, while conduct problems are associated with maternal maladjustment and with permissiveness and disciplinary ineffectuality on the part of fathers.

The specificity which these propositions imply was not evident in the present results. Instead, a single attitudinal pattern, consisting most prominently of paternal harshness, was related indiscriminately to all three of the problem dimensions examined. Of many possible interpretations, the most tenable have to do with age differences between the present groups of children and the ones examined previously. The behavior of young children is massive, and it is stimulus-bound. Whatever differentiation of problem behavior five-year-olds display may be less a function of intrapersonal tendencies, determined by specific kinds of parental attitude, than of a general personality disturbance, shown now in one form and now in another, as situational provocations and opportunities arise. From the study of older children, it seemed that paternal firmness was needed to prevent the occurrence of conduct problems in children. Limits are probably necessary all along, but, for these younger children, warmth, affection, and freedom from wrath seem more important in preventing problems of all kinds.

SUMMARY

Previous research on parent-child relationships led to three guiding hypotheses for the present study.

1. The attitudes of fathers are at least as intimately related as the attitudes of mothers to maladjustive tendencies among children.

2. Personality problems among children are related to harsh, autocratic attitudes and lack of parental concern among fathers.

3. Child conduct problems are related to maternal maladjustment and to permissiveness and disciplinary ineffectuality on the part of fathers.

Fifty-three kindergarten children (Kindergarten group) and 24 children attending a guidance clinic (Clinic group) participated in the investigation. Their parents were interviewed and rated on five factors of parental attitude defined in previous research by Sears, Maccoby, and Levin (6). A problem checklist applicable to children was factor analyzed, yielding three dimensions of problem behavior. Independent ratings of these were obtained from teachers and parents. Attitudes for Kindergarten and Clinic parents were then compared, and correlations established between parental attitudes and child problems.

The first hypothesis stated above was clearly supported, and the importance of paternal attitudes strongly reaffirmed. The differentiation implied by the second two hypotheses, however, was not apparent in the results. Instead, a single parent attitude pattern was diffusely associated with personality problems, conduct problems, and autism as well. The most conspicuous elements in this pattern were the strict, cold, aggressive attitudes of the fathers. Among the older children examined in the previous research, a considerable amount of paternal firmness seemed to be necessary for the prevention or amelioration of child conduct problems. For the younger children of the present study, love and kindness seemed more generally important.

REFERENCES

1. BECKER, W. C., PETERSON, D. R., HELLMER, L. A., SHOEMAKER, D. J., & QUAY, H. C. Factors in parental behavior and personality as related to problem behavior in children. *J. consult. Psychol.,* 1959, 23, 107-118.

2. EYSENCK, H. J. *The structure of human personality.* Wiley, 1953.

3. Himmelweit, H. T. A factorial study of "Children's behavior problems." Unpublished ms., Univer. of London, 1952.

4. Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika,* 1958, 23, 187-200.

5. Peterson, D. R., Becker, W. C., Hellmer, L. A., Shoemaker, D. J., & Quay, H. C. Parental attitudes and child adjustment. *Child Develpm.,* 1959, 30, 119-130.

6. Sears, R. R., Maccoby, E., & Levin, H. *Patterns of childrearing.* Row, Peterson, 1957.

DISCRIMINATION PERFORMANCE AS A BASELINE FOR INDIVIDUAL ANALYSIS OF YOUNG CHILDREN [1,2]

SIDNEY W. BIJOU*

University of Washington

It is well established that unequivocal relations between manipulated variables and instrumental or operant behavior may be demonstrated in single infrahuman organisms (4). The question raised here is whether such can be shown in research with young children. If reliable antecedent-consequent relations can be derived from individual analysis (the kind in which the subject serves as his own control), this type of design may be added to the investigator's repertory of skills with certain advantages. For example, individual analysis makes it possible to do considerably more research in situations in which only a limited number of subjects is available, it encourages functional analytic research with children with developmental anomolies, it makes possible controlled study of detailed and gradual developmental changes over long periods, and, because of its flexibility, it facilitates explorations into new procedures and functional relationships. These features and others have been discussed in some detail elsewhere (3).

Data from an individual preschool child will be presented consisting of baseline behavior and variations in performance following the introduction of a specific experimental condition. Two replications on two other young children will also be described.

The children were attending a private day-care center in Seattle and were observed in a mobile research laboratory parked adjacent to the school building.[3] The experimental room resembled a playroom and contained two tables with a toy on one and the apparatus on the other.

The toy, a standard commercial item (Playskool), consisted of a lever projecting from a wooden cage which contained a wooden ball. Striking

---

* Gatzert Institute of Child Development, Department of Psychology, University of Washington, Seattle 5.

[1] This paper with minor modifications was presented at a symposium, "Recent developments in experimental methods with children," at the annual meeting of the Amer. Psychol. Ass., Washington, D.C., September, 1958.

[2] This research was supported by grants from the National Science Foundation (G2141) and the National Institute of Mental Health (M-2208), U. S. Public Health Service.

[3] We wish to express our gratitude to Mrs. Charlotte Pearson and her staff of the Carrousel Nursery School for their exceptionally fine cooperation. We also wish to thank Robert Sharpley and Persis T. Sturges for assisting in the study.

Child Develpm., 1961, 32, 163-170.

the lever flipped the ball to the top of the cage. As it fell, it rattled to its original position and in a position to be tossed up again by the next response. The child could play with the toy at any time.

The apparatus was a colorfully painted box with two jewel lights, a metal handle protruding from the front panel, and a tray on one side for catching reinforcements falling from a chute extending through an opening in the wall separating the experimental and control rooms. Pressing the handle down (820 grams of pressure were required to make a complete response) was always followed by a relay click, and now and then by a reinforcer dispensed by a Gerbrands Universal Feeder in the control room. Responses are recorded on a Gerbrands cumulative recorder. A young lady (referred to as *A* in distinction to *E*, the experimenter) brought the child into the room, remained with him during the session, and returned him to school. She followed a routine for removing his coat, taking him to the center of the room, giving him instruction ("Here are some toys. You may play with them as long as you like."), and retiring behind a portable partition to read a book and carry out variables in procedures suggested by *E*. *E* was behind a one-way screen in the wall opposite to that where *A* was stationed. In addition to observing the child (*A* did not participate in observing the subjects), he monitored the apparatus and talked to *A* through a telephonic system which delivered signals audible only to *A*. Detailed descriptions of the mobile laboratory, equipment, and procedures have been published (1, 2).

Training was given to establish a discrimination marked by two minutes of rapid responding in the presence of an amber light on the left side of the panel and by two minutes of no responding to a blue light on the right side. When the amber light was on, reinforcements (trinkets, dime-sized cookies, sugar corn pops, and M & M candy, mixed in respective proportions of 5:3:3:1) were distributed on a variable ratio ranging from 10 to 90 with a mean of 50. Sessions were 20 minutes and intervals between sessions six days. Because of absences, the interval was extended to seven to eight days at times. After two sessions of stable performance, a toggle switch was installed over each light to enable the subject to turn out the lights.[4] Changes in discriminative behavior as a function of this alteration were observed.

Data from the first child, JDA, a boy, 4 years, 3 months of age, are shown in Figure 1 and the upper part of Table 1.

Chart A is a record of his seventh day and the second session in which stable discrimination was shown. Here, as well as in the other charts, the lower horizontal line is a time line which indicates occurrence of certain events, including change in the discriminative stimuli. When the line is down, the amber light is on for two minutes, and, when it is up, the amber light is off and the blue light is on for two minutes. The graph is a record of responses. When the apparatus handle is pressed all the way down, the

---

[4] To give the child a brief opportunity to explore the new attachments, working the switches during the first two minutes did not turn off the lights.

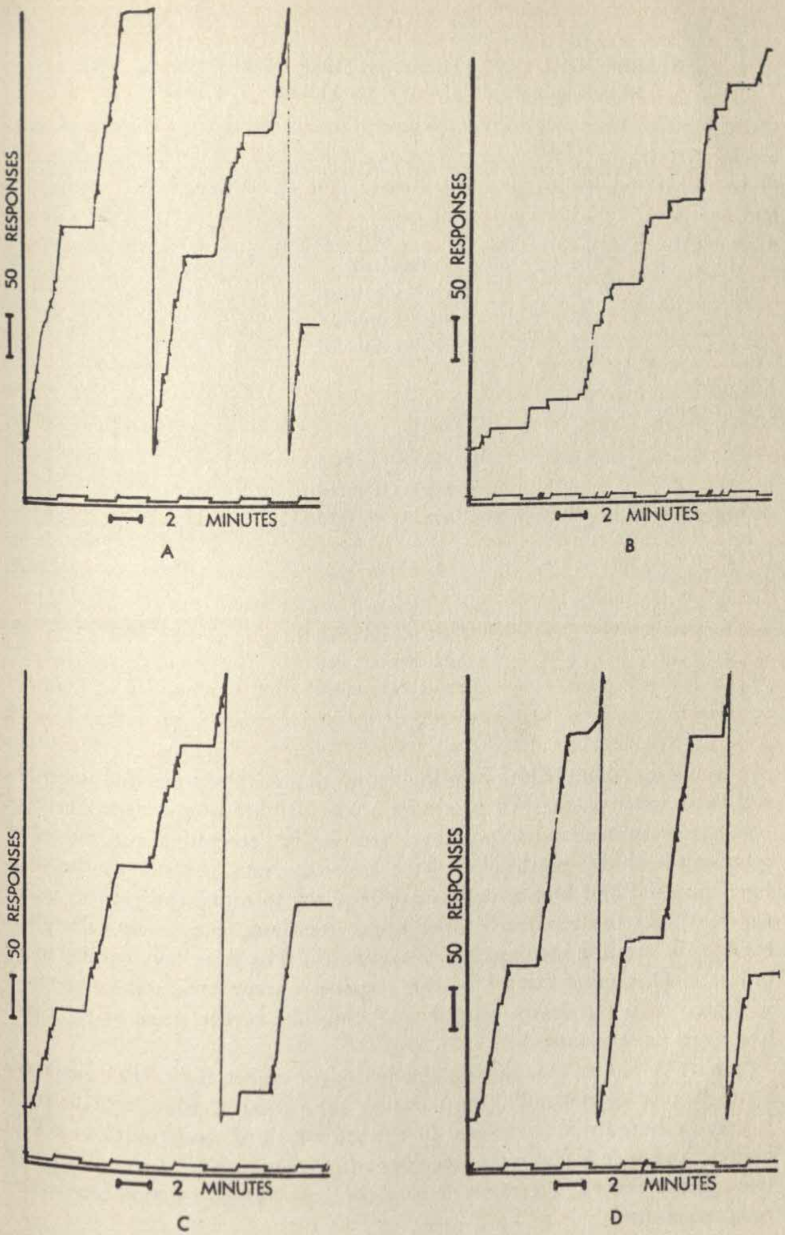

FIGURE 1—Performance of a 4-year, 3-month-old boy (JDA) on the baseline day (chart A) and three succeeding days (charts B, C, and D) after introduction of a new experimental condition.

TABLE 1

EXPERIMENTAL CONDITIONS, NUMBER OF RESPONSES, AND REINFORCEMENT SHOWN IN FIGURES 1, 2, AND 3

Session	*Days Since Previous Session*	*Experimental Condition*	*No. of Responses*	*No. of Reinforcements*
JDA, Boy 4 yrs., 3 mos.				
7	7	Standard	1260	25
8	6	Switches installed	581	13
9	6	Switches installed	852	16
10	8	Switches installed	1265	25
KBB, Girl 3 yrs., 10 mos.				
9	6	Standard	1590	32
10	6	Switches installed	1231	24
11	6	Switches installed	936	19
12	6	Switches installed	1094	21
PMC, Girl 4 yrs., 10 mos.				
9	6	Standard	1633	25
10	6	Switches installed	578	20
11	6	Switches installed	654	20
12	7	Switches installed	730	14

pen moves one-hundredth of an inch upward from the time line; when the handle is inactive, the pen produces a line parallel to the time line. The result is a cumulative curve. Pauses are readily perceptible and rate of responding is shown by the slope—the faster the rate, the steeper the slope. Five hundred and fifty responses are required to move the pen across the paper. When the pen reaches the upper terminus, it automatically glides back to the starting position. Resets are indicated by thin lines parallel to the ordinate. One other feature of the response curve: each reinforcement is indicated by a pip drawn roughly perpendicular to the slope of the curve, like a pip on the time line.

Chart A shows that during the period of amber light JDA maintains a steady rate of responding and during the period of blue light he stops. It may be noted that at the end of the second, third, and fourth blue light periods he made a few responses suggesting that he had discriminated the two-minute period. He made a total of 1260 responses and received 25 reinforcements.

Prior to session 8, the toggle switches were installed. JDA was brought into the room in the usual manner and given the usual instruction: "You may play with the toys as long as you like." As shown in chart B, his per-

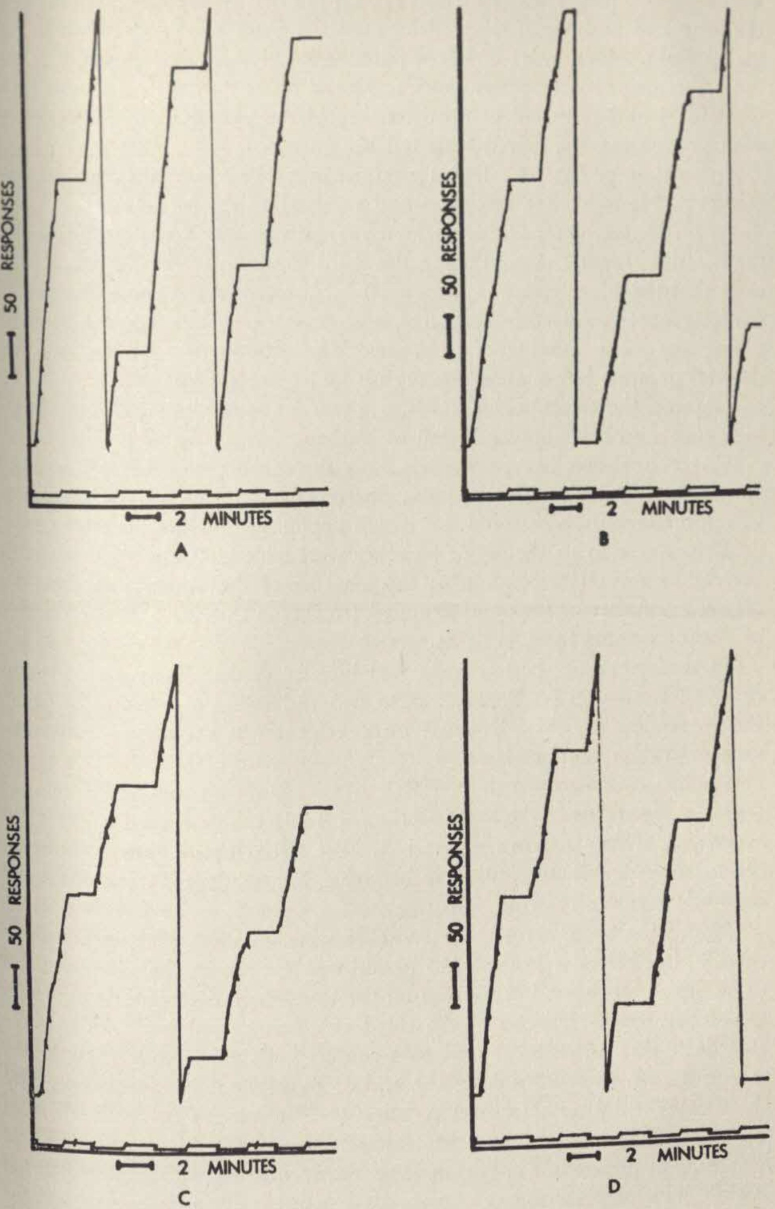

Figure 2—Performance of a 3-year, 10-month-old girl (KBB) on the baseline day (chart A) and three succeeding days (charts B, C, and D) after introduction of a new experimental condition.

formance was much poorer than in session 7. (Pips on the time line indicate the time and number of relay clicks made by *E* to call *S*'s attention to the amber light.) One index of the difference between session 7 and 8 is the decrease from 1260 to 581 responses. Several times during the session JDA went to *A* and asked about the meaning of the switches. In conformance with procedures, she merely reflected his questions.

In session 9, chart C, JDA's performance was more efficient than in session 8. He made 852 responses and received 16 reinforcements.

In session 10, chart D, his performance was similar to that in session 7, the baseline day. (The pip during the third elevation of the time line indicates that the blue light was turned off.) He received the same number of reinforcements received in session 7, and he came within five responses of giving the same number of responses. The introduction of the switches showed greatest influence in the eighth and ninth sessions.

Data on the second child, KBB, a 3-year, 10-month-old girl, are shown in Figure 2 and the middle section of Table 1.

Chart A shows her performance on the ninth experimental day, the second day of stable discrimination. She made 1590 responses and received 32 reinforcements, a performance which approaches maximum efficiency.

Before session 10, the toggle switches were added. As shown in chart B, the child's rate of responding in the presence of the amber light was decreased. Number of responses dropped from 1590 to 1231 and the number of reinforcements from 32 to 24.

In session 11 her performance was inferior to that in 10 as shown in chart C. Responses fell from 1231 to 936 and reinforcers from 24 to 19. But in session 12, chart D, some improvement was shown with 1094 responses and 21 reinforcers.

Findings on the third child, PMC, a girl 4 years, 10 months of age, are given in Figure 3 and the lower section of Table 1. The record of the ninth experimental day is shown in chart A. Like KBB, it took nine sessions for her to reach stable discriminative behavior. She made 1633 responses and received 25 reinforcements.

Before the tenth session, the switches were installed, and, as shown in chart B, her behavior in the tenth period was less precise than in the ninth. (The pip in the second depression on the time line indicates time-out to go to the bathroom. The one in the third elevation period indicates that the blue light was turned off.) Due to a conflict with the school schedule, the session had to be terminated at the end of 18 minutes. In comparison with the first 18 minutes of session 9, number of responses emitted fell from 1621 to 578 and number of reinforcements received from 25 to 20. There was some improvement in session 11 as shown in chart C. Variability in the initial phase of performance is correlated with turning off the amber light on the first and second presentations. In session 12, chart D, her responding continued to be uneven as compared with session 9, chart A. (She turned the amber light off on the first and second appearances and the blue off on the

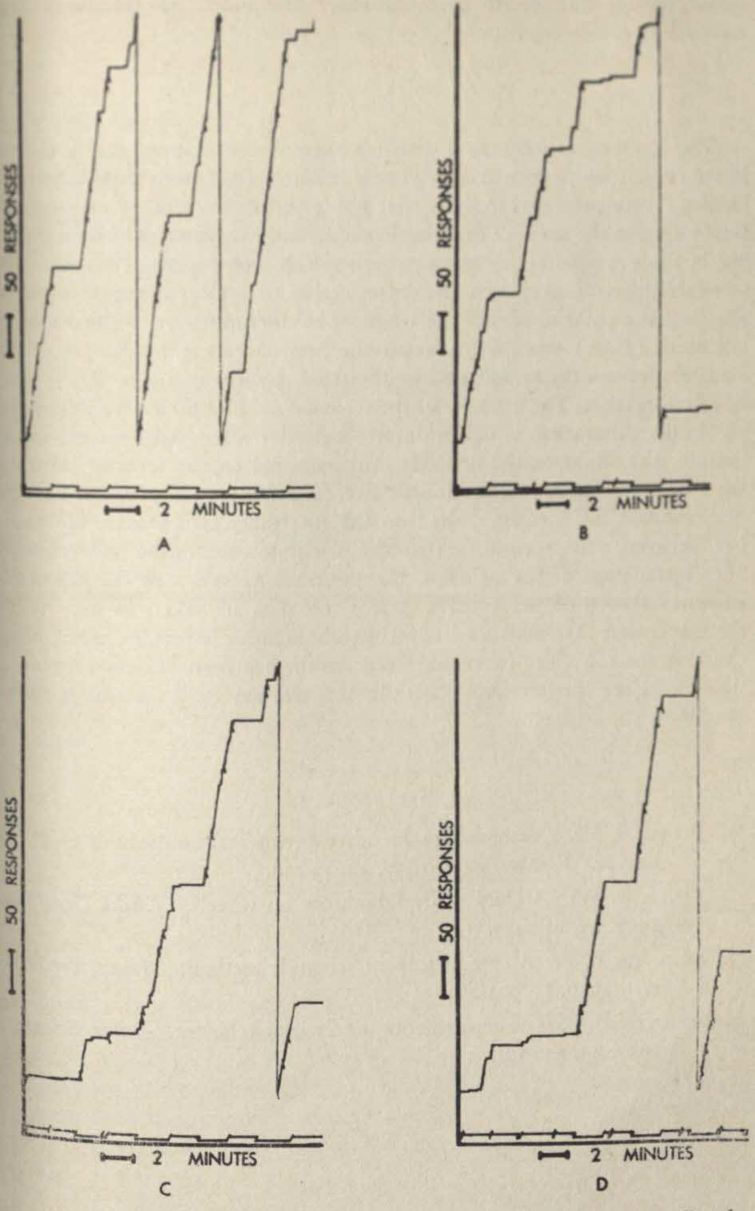

Figure 3—Performance of a 4-year, 10-month-old girl (PMC) on the baseline day (chart A) and three succeeding days (charts B, C, and D) after introduction of a new experimental condition.

second, third, and fourth presentations.) She made 730 responses and received 14 reinforcements.

SUMMARY

The question raised here is whether unequivocal relations can be shown in the *individual* performance of young children. Data from three preschool children were presented to show that stable performance of two minutes of responding in the presence of an amber light and two minutes of no responding in the presence of a blue (a pattern which took from six to eight days to establish) was altered by the introduction of a pair of toggle switches allowing the child to control the lights. The discriminative performance of one child (JDA) was most affected the first day after installation of the switches, less on the second, and by the third day it was similar to the baseline performance. The behavior of the second child (KBB) showed the opposite trend. Alteration in discriminative behavior after experimental modification was slight on the first day and increased on the second and third days. Recovery began on the fourth day. The third child (PMC), like the first, showed the greatest change on the first experimental day. However, her "recovery rate" was slower than the first child's for, in the fourth session after installation of the switches, her performance was still far below the efficiency shown on her baseline day. These data are taken as support for the contention that individual experimental analysis is feasible with young children since a clear functional relationship has been shown between a stable baseline performance and the introduction of a special stimulus condition.

REFERENCES

1. BIJOU, S. W. Methodology for an experimental analysis of child behavior. *Psychol. Rep.,* 1957, 3, 243-250.
2. BIJOU, S. W. A child study laboratory on wheels. *Child Develpm.,* 1958, 29, 425-427.
3. SKINNER, B. F. A case history in scientific method. *Amer. Psychologist,* 1956, 11, 221-233.
4. SKINNER, B. F. The experimental analysis of behavior. *Amer. Scientist,* 1957, 45, 343-371.

CULTURAL STEREOTYPES AND THE VALIDITY OF INTERVIEWS FOR RESEARCH IN CHILD DEVELOPMENT[1]

JOAN McCORD *and* WILLIAM McCORD*
Stanford University

One of the fundamental problems in the study of child development is the difficulty in obtaining valid information regarding childrearing practices. Most studies of child development are based upon interviews with the parents. Yet the interview as a technique for analyzing parent-child or mother-father relations has been subjected to serious criticism. After a comparison of the results of two pieces of similar research on social class differences in approaches to child training which yielded different results, Havighurst and Davis were forced to conclude: ". . . this reinforces the conviction of the present writers that the interview is not nearly so good as participant observation for securing data both on the behavior and the attitudes of parents toward their children" (1, p. 442). In their recent, extensive study of child development, Sears, Maccoby, and Levin (despite the care with which the interviews were conducted) had to remind their readers of the possibility that some of their findings "may have resulted from a response set created by the interview itself" (6, p. 63). And, in an article summarizing research on interview techniques, Eleanor and Nathan Maccoby have pointed out: "Validity studies all suggest that when people are interviewed directly concerning behavior about which there is a strong expectation of social approval or disapproval, and in which there is considerable ego-involvement, respondents tend to err in the direction of idealizing their behavior" (2, p. 482). Although researchers agree on the possible dangers inherent in interviewing, there has been little research directed toward the identification of the specific areas in which considerable ego-involvement can be expected.

---

* Sociology Department, Stanford University, Stanford, California.

[1] The authors wish to express their gratitude to the Ella Lyman Cabot Foundation, the Harvard Laboratory of Social Relations, and the National Institute of Mental Health (Grant M-2647) for their generous financial support of this research.

Child Develpm., 1961, 32, 171-185.

Even though most researchers are agreed that important biases may mar the results of studies based on interviews, practical alternatives are by no means easy to find. In this article, we will attempt to suggest those areas of childrearing in which the interview may be an adequate technique, as well as indicating certain systematic interview biases which are apparently due to cultural values or to strong expectations of social approval or disapproval.

As many researchers have commented, for an ideal study of childrearing, one would like to place an invisible, unbiased observer in the family for a period of several years. The Cambridge-Somerville Youth Study provided what is probably the closest approximation to such an invisible, unbiased observer. Begun in 1935, under the guidance of Richard Clark Cabot, this project was designed primarily for the purpose of preventing delinquency and of testing the effectiveness of a preventive program. Although the attempt to deter delinquency seems to have failed, the project resulted in a unique fund of information regarding childhood behavior, family interaction, and the transmission of values. Five hundred ten lower- and lower-middle-class boys were included in the study. These boys were divided into a "treatment" and a "control" group after data had been obtained from medical and psychological examinations and reports from home visitors who interviewed the mother regarding her childrearing practices. On the basis of this information, each boy was matched to another as closely similar in background and in personality as was possible. One in each pair was then selected, by toss of a coin, to receive the services of the Youth Study. The treatment program began in 1939 and continued until 1945. The "control boys" and their families were interviewed periodically during these years. The division between treatment and control groups and the attendant difference in available information about childrearing practices provided the basis for the present study. Those boys and their families who received treatment (hereafter called the "observation group") were directly observed by several workers over a period of more than five years. The control group (hereafter called the "interview group") were interviewed several times in the standard fashion. Thus, the information which emerged from this study was derived from fundamentally different sources.

METHOD

The Observation Group

For the boys who received treatment, complete records were reported after each contact by the counselor. Counselors gathered information from teachers, ministers, parents, and neighbors detailing the backgrounds of each of their boys. More importantly, the counselors repeatedly visited the boys and their families. The families were visited usually weekly or bimonthly, and visits continued for a median of five and three-quarters years. The counselor-observers (except for one who was a nurse) were trained social workers, psychologists, or youth workers. They described in detail whatever activities they observed in the home, at school, during chance meetings on the streets, and in the Cambridge-Somerville Youth Study center. They reported conversations with parents, friends, neighbors, and, of course, the subjects themselves.

In 1957, these raw observations were coded by trained researchers. Rigid definitions of the categories and repeated tests of interrater agreement resulted in a relatively high degree of reliability and, presumably, the elimination of "the halo effect" in rating.

The Interview Group

For the control boys, records had been kept of the original interviews with the mothers which had been used for matching cases. These were semi-structured. Although the exact order and specific wording of the questions were left unspecified, the interviewers were to fill out information sheets including:

> A. History of subject—eating habits (What good eating habits does subjects have? What poor habits?), sleeping habits (How long does he sleep? Do you have trouble getting him to bed?), recreational habits (What does subject like to do?), attitude toward school (Do you have much trouble with subject about truanting? What does he like best about school? What does he like least?), attitude toward discipline (What does subject do when you have to punish him?), attitude toward parents (How does subject feel about his father? Give an example. How is he with you?), personality (What kind of boy is subject? How do the neighbors get along with him?).
>
> B. Father's history—occupational background, education, personality (What is your husband like? Do you have trouble getting him to do things you want?), methods of discipline (How does he get subject to do what he should? What happens if subject is bad? Can you give an example?), attitude toward subject (How does your husband get along with subject? Does he have plans for subject's future? Do they play together often?)
>
> C. Mother's history—same as for father.

In addition, at least one interview with the boy and three unstructured interviews with the mother had been gathered for each case. These latter pertained primarily to questions of disciplining the child, the mother's attitude toward the subject and his siblings (if any), and a general picture of family life.

The interview records, too, were categorized with high interrater agreement.

Since the same individuals gathered the records on the observation group and the interview group, it seems reasonable to assume that, if their attitudes affected the nature of the data recorded, these errors would be similar for the two groups.

These two types of records—those gleaned from continuous reports of direct observation over a number of years and those summarizing several intermittent interviews with the mother—furnish the data for the present analysis.

We reasoned that if the observation records were valid and if the interviewed families were actually similar to the observed families in behavior and attitudes, any apparent differences in reported information might be correctly attributed to systematic interview biases. Therefore, before turning to the results of this comparison, we must consider two questions: (a) evidence for the validity of the information gathered about the observation group, and (b) evidence that the childrearing practices of the interview group were in fact similar to those in the observation group.

The Problem of Validity

The face validity of the observation records is considerable. The families had been visited repeatedly by trained observers, the observations had been recorded almost immediately after each visit, and the families were not aware of the fact that records were being kept. (As far as the parents were concerned, the Cambridge-Somerville Youth Study was a "service" organization.)

Counselor turnover, considered a potential disadvantage for treatment, proved an advantage for research: the vast majority of subjects were observed by two to four counselors. Thus, possible individual observer biases tended to cancel one another.

Perhaps the most important evidence for the validity of the observation records is the highly significant relations yielded by these records between childrearing practices and the criminality, alcoholism, and psychoses of the boys *as adults* (3, 4, 5). Thus, we know that the observation records have good "predictive power."

TABLE 1

ECOLOGICAL BACKGROUND OF SUBJECTS

	Observation Group ($N = 255$)	*Interview Group* ($N = 255$)
	PERCENTAGE	
Race		
White	91	91
Negro	9	8
Mixed	0	1
Father's Birthplace		
United States	49	44
Foreign-born	49	52
Not ascertained	2	4
Neighborhood in Childhood		
Residential	37	35
Transitional	62	62
Not ascertained	1	3

TABLE 2

CULTURAL BACKGROUND AND AGE OF PARENTS

	Observation Group (*N* = 255)	*Interview Group* (*N* = 255)
	PERCENTAGE	
MOTHER		
Religion		
Catholic	62	57
Protestant	36	38
Other	0	2
Not ascertained	2	3
Education		
College graduate	1	0
High school graduate	10	8
Some high school	15	13
Grammar school graduate	17	20
Some grammar school	33	32
No formal education	11	9
Not ascertained	13	18
Age of Parent When Boy Was Born		
Under 20	8	6
20 - 24	21	19
25 - 29	21	22
30 - 34	25	19
35 - 39	13	14
40 - 44	4	7
Over 44	1	1
Not ascertained	7	12
FATHER		
Religion		
Catholic	61	54
Protestant	30	34
Other	2	2
Not ascertained	7	10
Education		
College graduate	1	1
High school graduate	11	9
Some high school	12	12
Grammar school graduate	19	19
Some grammar school	31	29
No formal education	13	9
Not ascertained	13	21
Age of Parent When Boy Was Born		
Under 20	0	0
20 - 24	11	11
25 - 29	23	20
30 - 34	24	22
35 - 39	17	18
40 - 44	11	13
Over 44	7	5
Not ascertained	7	11

The Question of Comparability

Three types of evidence indicate that the observation group and the interview group were very similar in attitudes and behavior.

First, the two groups gave similar responses when interviewed at the beginning of the project. In fact, the two groups were matched largely on the basis of information given in these interviews. Although systematic biases tend to interfere with the validity of information gathered by interviews and are probably related to actual childrearing techniques, the fact that the two groups responded in a similar fashion indicates that similar processes controlled the nature of these responses.

Secondly, we compared the two groups using "objective" variables which were least likely to be subject to bias and yet might be correlated with differences in childrearing behavior. The figures are presented in the following charts.

We found no significant differences[2] between the two groups when we compared their racial background, their father's birthplace, or the neighborhoods in which the boys had been reared (*see* Table 1).

We found practically no differences between the groups in the cultural background or age of their parents (*see* Table 2).

The similarities between the groups regarding education of the father and place of residence point to equal social status between the groups. Nevertheless, we found a slight, but statistically nonsignificant ($\chi^2 = 2.80$), difference in the occupational status of the boys' fathers. This difference, we believe, might be attributable to misleading statements of the interview group since it is in a "culturally approved" direction (*see* Table 3).

The striking similarities in family background found when we compared race, neighborhood, father's birthplace, religion, education, and parental age lead one to believe that the two groups closely resembled each other in terms of objective social characteristics.

We considered the possibility that the number of children in the family, the ordinal position of the child, and the sex of his siblings might affect childrearing practices. Again, we found that the observation and interview groups had essentially similar compositions (*see* Table 4).

Thus, by relatively objective criteria, the observation group and the interview group were basically similar; one would presume, therefore, that the two groups would also be similar in childrearing techniques.

Our third approach to ascertaining the comparability of the two groups was based upon the relationships we had found between childrearing practices and adult social deviance (3, 4, 5). We reasoned that, if adult social deviance is strongly related to the behavioral observations made in childhood, then a similarity in adult deviance patterns should indicate a similarity in childrearing techniques. The incidence of adult deviance of the boys (as

---

[2] Tests of significance were two-tailed, using $p < .05$ as the minimum standard for asserting significance.

TABLE 3

FATHER'S OCCUPATION

	Observation Group (*N* = 255)	*Interview Group* (*N* = 255)
	PERCENTAGE	
Professional	1	1
White collar	8	12
Skilled tradesman	29	29
Unskilled worker	62	58

TABLE 4

FAMILY COMPOSITION OF SUBJECTS

	Observation Group (*N* = 255)	*Interview Group* (*N* = 255)
	PERCENTAGE	
Number of Children in Family		
One	5	5
Two	13	12
Three	15	14
Four	15	16
More than four	52	53
*Ordinal Position of Boy**		
Only child	5	6
Twin	2	2
Oldest	17	15
Middle	48	49
Youngest	28	28
Sex of Siblings		
Male only	14	12
Female only	13	13
Male and female	66	67
None or twin	7	8

* A boy was rated as an only child if no other siblings were living in the home.

TABLE 5

RECORD OF SOCIAL DEVIANCE IN ADULTHOOD*

	Observation Group	*Interview Group*
	NUMBER OF SUBJECTS	
Noncriminal alcoholics	11	10
Nonalcoholic criminals	44	44
Alcoholic criminals	18	12
	73	66

* Subjects were considered alcoholics if they had been members of A.A., received medical treatment for alcoholism, or had been convicted more than once for public drunkenness; they were considered criminals if they had been convicted by the courts for theft, assault, or sex crime at least once.

ascertained through perusal of community records when the boys reached adulthood) was almost identical for the two groups (*see* Table 5).

In view of the striking similarities between the two groups on all of these measures, we are reasonably confident that any major differences in reports on childrearing practices between the two groups as reported through interviews resulted from biases inherent in the interviews, rather than actual differences. On the other hand, for those variables on which we found little variation between the reports of the participant observers and the interviewers, there is indirect evidence that the interviews produced valid information.

RESULTS

In analyzing the similarities and differences, we will consider three phases of family life which affect childrearing: first, the emotional tone of the family; second, leadership and the division of responsibility between the parents; and third, techniques used for controlling the child.

The Emotional Tone of the Family

For the observation group, of course, actual behavior was observed over an extended period of time. For the interview group, ratings were based on summary judgments of a minimum of four interviews. Measures were designed to record the emotional relation between the mother and child, the father and child, and between the father and mother. *Some, but not all, of these measures seemed to be subject to interview bias.*

When we looked at the comparative ratings of the parents' attitudes toward their sons, we found that for both mothers and fathers *significantly less active rejection was revealed through the interviews* (*see* Table 6).

TABLE 6

PARENTAL ATTITUDES TOWARD THEIR SONS*

	MOTHER		FATHER	
	Observation Group ($N = 250$)	*Interview Group* ($N = 241$)	*Observation Group* ($N = 216$)	*Interview Group* ($N = 210$)
	PERCENTAGE		PERCENTAGE	
Warm	50	46	30	32
Passive	32	44	46	55
Rejecting	18	10	24	13
	$\chi^2 = 8.1$; $p < .005$		$\chi^2 = 8.0$; $p < .005$	

* "Warm" parents evidenced active affection for their children through open approval and pleasure in being with them; "passive" parents had little interaction with their sons although they may have been concerned for their welfare; "rejecting" parents openly demonstrated their dislike for their offspring. Agreement of two independent raters on a random sample of the observation group was 83 per cent for the mothers' and 77 per cent for the fathers' attitudes; on a random sample of the interview group, agreement was 84 per cent for the mothers' and 76 per cent for the fathers' attitudes.

Apparently, even repeated interviews failed to differentiate passive parents from those who actively rejected their children. Yet there seems to have been little systematic error in distinguishing the actively affectionate parents from those who were *either* passive or rejecting.[3]

Not surprisingly, we found that *the interviews indicated significantly fewer negative attitudes of the boys toward their parents* (*see* Table 7). Sixteen per cent of the total 255 observed subjects had revealed disapproving or fearful attitudes toward their mothers and 33 per cent had evidenced dislike of their fathers. In the interview group, however, only 7 per cent gave evidence of negative attitudes toward their mothers and 19 per cent toward their fathers.

Despite the inadequacies of the interview in detecting rejection between parent and child, there do seem to be indices of the family's emotional tone which are relatively free from systematic bias. There was very little difference between the observation and the interview groups either on the measure of parental attitudes toward each other or on the amount of conflict between the parents. Approximately one-fourth of each group of parents were rated as antagonistic and approximately one-third evidenced intense conflict (*see* Table 8).

Thus, in terms of emotional interaction, the interview appears to be more reliable as a technique for judging the attitudes of parents toward each

---

[3] Hence, we would hazard the suggestion that scales of attitudes of parents toward their children be based on positive signs of warmth rather than upon signs of rejection.

TABLE 7

BOYS' ATTITUDES TOWARD THEIR PARENTS*

	ATTITUDE TOWARD MOTHER		ATTITUDE TOWARD FATHER	
	Observation Group ($N = 220$)	*Interview Group* ($N = 208$)	*Observation Group* ($N = 203$)	*Interview Group* ($N = 187$)
	PERCENTAGE		PERCENTAGE	
Favorable	82	91	59	74
Unfavorable	18	9	41	26
	$\chi^2 = 7.4$; $p < .01$		$\chi^2 = 10.7$; $p < .005$	

* The boy's attitude toward his parent was considered "favorable" if there was some evidence of general approval and "unfavorable" if the boy appeared to fear, disapprove, or disdain his parent. Interrater agreement for the observation group was 93 per cent for attitude toward mother and 80 per cent toward father. For the interview group, interrater agreement was 88 per cent toward mother and 96 per cent toward father.

TABLE 8

PARENTAL WARMTH AND CONFLICT*

RELATION BETWEEN PARENTS			DEGREE OF PARENTAL CONFLICT		
	Observation Group ($N = 205$)	*Interview Group* ($N = 182$)		*Observation Group* ($N = 214$)	*Interview Group* ($N = 207$)
	PERCENTAGE			PERCENTAGE	
Affectionate	60	62	Little	30	34
Indifferent	12	9	Some	33	35
Antagonistic	28	29	Considerable	37	31

* Parents were considered "affectionate" if there was evidence of mutual concern for one another's welfare, "indifferent" if the parents seemed unconcerned about each other, and "antagonistic" if they seemed to take pleasure in discomfitting one another. Interrater agreement was 73 per cent for the observation group and 80 per cent for the interview group. Interrater agreement regarding amount of parental conflict was 73 per cent in the observation group and 68 per cent in the interview group.

other than of attitudes between parents and their children. These findings suggest that the admission of parental conflict and antagonism is more "permissible" than is a confession of parent-child antagonisms.

Leadership and Division of Responsibility

Who makes the decisions in the family? Who carries them out? The answers to these questions impinge upon the child's perception of the world and, to a considerable extent, have been found to be related to his adult

TABLE 9

ROLE DIFFERENTIATION*

	Observation Group ($N = 203$)	*Interview Group* ($N = 205$)
	PERCENTAGE	
Highly differentiated	49	60
Not highly differentiated	51	40
	$\chi^2 = 5.7$; $p < .02$	

* Interrater agreement was 70 per cent for the observation group and 92 per cent for the interview group.

behavior (3, 4, 5). In assessing the answers to these questions, we found several apparent interviewing biases.

One aspect of the division of responsibility pertains to role differentiation. Do mother and father share in such activities as caring for the children, earning a livelihood, preparing meals, cleaning house? Considering only those families in which both mother and father were available as models, we found that *the interview reports indicated a greater degree of differentiation than was found among the observation group* (*see* Table 9).

Parental dominance was a measure designed to show which parent made important decisions for the family (such as where to go on vacations, how to spend money, what to do when crises arose). Judgments based on interviews differed significantly from those based on direct observation: *The interviewed mothers tended to portray their husbands as more dominant* (*see* Table 10). These differences between the interview and the observation

TABLE 10

PARENTAL DOMINANCE*

	Observation Group ($N = 254$)	*Interview Group* ($N = 228$)
	PERCENTAGE	
Father	39	49
Mother	32	25
	$\chi^2 = 5.2$; $p < .025$	
Equal	7	7
Only one parent at home	22	19

* Interrater agreement was 83 per cent for the observation group and 84 per cent for the interview group.

TABLE 11

PARENTAL ESTEEM*

	Esteem of Father for Mother		Esteem of Mother for Father	
	Observation Group ($N = 186$)	*Interview Group* ($N = 165$)	*Observation Group* ($N = 223$)	*Interview Group* ($N = 199$)
	PERCENTAGE		PERCENTAGE	
Moderate or high	63	71	54	66
Low	37	29	46	34
	$\chi^2 = 2.7$; *ns*		$\chi^2 = 6.9$; $p < .01$	

* Interrater agreement of esteem of father for mother was 80 per cent for the observation group and 96 per cent for the interview group. Agreement between raters on esteem of mother for father was 90 per cent on the observation group and 92 per cent on the interview group.

groups illustrate the American emphasis on differentiating the role of husband and wife, with the father occupying the role of family leader.

Further indication of the tendency to portray the father in the role of leader was found in the analysis of reports on the amount of respect the parents had for one another (*see* Table 11). One would expect that, if interview bias was due primarily to a direct attempt to give oneself status, there would be considerable discrepancy between interviews and observations in the measure of the father's esteem for the mother—i.e., one would expect the mothers to indicate that their husbands held *them* in high esteem. Although there was some tendency in this direction, the difference between the interview and the observation records was not statistically significant. On the other hand, there did appear to be some interviewing bias in measuring the mother's esteem for the father. Thus, although there was a slight tendency for the mothers to overvalue themselves, there was a much stronger tendency for them to indicate higher esteem of their husbands—in other words, their interviews presumably reflected a cultural stereotype of the wife "respecting" her husband.

The same phenomenon occurred in the comparative analysis of the "leadership roles" of the parents. For this measurement, each parent was considered individually (e.g., both parents might be "leaders" although one was dominant); the interviewed mothers more often pictured their husbands as "dictators" or "leaders" (*see* Table 12). It should be remembered that judgments for the observation group were based on reports of the actual interaction among family members and that judgments for the interview group were based on at least four interviews with the mother and not simply on their responses to one question. Consciously or unconsciously, therefore, the interviewed mothers must have wanted to create the general impression

TABLE 12

PARENTAL ROLES*

	MOTHER		FATHER	
	Observation Group ($N = 247$)	*Interview Group* ($N = 231$)	*Observation Group* ($N = 213$)	*Interview Group* ($N = 203$)
	PERCENTAGE		PERCENTAGE	
Dictator	4	5	14	17
Leader	65	63	45	54
Passive	15	15	36	24
	$\chi^2 = 0.7$; *ns*[†]		$\chi^2 = 7.3$; $p < .05$[†]	
Ambiguous	16	17	5	5

[*] Parents were considered "dictators" if they made all or most decisions without consulting one another and with almost complete authority. They were considered "leaders" if they accepted responsibility for many decisions. They were considered "passive" if they made few decisions; and their role was considered "ambiguous" if they were rarely involved in family affairs (e.g., if they came home only periodically) or if they made decisions and yet refused to take responsibility for them. Interrater agreement on mother's role was 73 per cent for the observation group and 76 per cent for the interview group; for father's role, interrater agreement was 73 per cent for the observation group and 64 per cent for the interview group.

[†] $df = 2$.

that their husbands were leaders in the home.[4] This, we can assume, reflects a generalized ideal for lower- and lower-middle-class American families.

Techniques of Control

There are, of course, many facets to the technique of control by which the parent transmits values to the child. There was no difference between the two groups in reports on degree of supervision exercised by the parents, in their reports as to which parent took major responsibility for discipline, or in judgments concerning the mother's method of discipline. The interviewed mothers did, however, tend to report their husbands as less punitive (*see* Table 13).

The greater bias found regarding the father's technique is open to several interpretations. Two seem particularly plausible. First, the results might be further evidence of the tendency of the mothers to "upgrade" their husbands. Second, the results might indicate a stereotype of ideal parenthood in which the mother, "because she must," is permitted to utilize physical punishments, but the father is expected to control his children through the authority of his position as "head of the home."

---

[4] Apparently there is less cultural stereotype regarding discipline. In both groups, approximately 60 per cent of the mothers were regarded as the major disciplinarian.

TABLE 13

TECHNIQUE OF DISCIPLINE*

	MOTHER		FATHER	
	Observation Group ($N = 247$)	*Interview Group* ($N = 218$)	*Observation Group* ($N = 206$)	*Interview Group* ($N = 197$)
	PERCENTAGE		PERCENTAGE	
Punitive	49	45	62	52
Nonpunitive	51	55	38	48
	$\chi^2 = 0.1$; *ns*		$\chi^2 = 4.4$; $p < .05$	

* Severe spankings and harsh verbal threats were considered as "punitive." Interrater agreement for mother's discipline was 97 per cent on the observation group and 92 per cent on the interview group; for father's discipline, interrater agreement was 77 per cent on the observation group and 84 per cent on the interview group.

SUMMARY

In an attempt to assess the validity of interviews in the investigation of child development, two groups of 255 families were compared. Information about one of these groups was gathered through extensive observation over an average of five and three-fourths years; information about the other group was obtained through a series of interviews with the mother in the home. Originally, the two sets of families had been matched to insure their comparability; analysis revealed that the "observation group" and the "interview group" were essentially similar in race, birthplace of father, neighborhood, religion, education of parents, age of parents, number of children in family, ordinal position of children, and sex of children. Thus, in terms of "objective" criteria, there were no major differences between the two sets of families.

We assumed, therefore, that any contrasts in reports on childrearing techniques resulted from interview biases. Analysis revealed that the interview group more often reported that: the mother did not actively reject her son ($p < .005$); the father did not actively reject his son ($p < .005$); the boy had a favorable attitude toward his mother ($p < .01$); the boy had a favorable attitude toward his father ($p < .005$); roles within the family were clearly differentiated ($p < .02$); the father was the dominant person in the family ($p < .025$); the mother regarded the father with high esteem ($p < .01$); the father played the role of "leader" for the family ($p < .05$); the father was nonpunitive in disciplining the son ($p < .05$).

On the other hand, there were no major differences in reports on such variables as: the parents' affection for each other, the parents' conflict with one another, the esteem of the father for the mother, the degree of supervision exerted by the parents over the child, the mother's technique of discipline.

It would appear, therefore, that the validity of the interviews was marred by the parents' tendency to make their picture of family life conform to cultural stereotypes. Specifically, it seems that the interviewed parents viewed the "ideal" family as one in which the parents loved their children (and vice versa), the father directed family affairs, and the father was a "decent" person who treated his son with kindness. Apparently, the interviewed mothers were more willing to place themselves in an honest, if unflattering, light than they were to undermine the "respectability" of their husbands. Thus, for a variety of possible reasons—perhaps because prolonged contact between the parents and the observers led to a greater willingness for the parents to reveal themselves, perhaps because the parents found it impossible to keep up a facade throughout the five-year period—it would seem that the information derived from direct observation of the families offered a more valid picture of their actual practices.

It is possible that changes in parent-child relations over a five-year period might account for some of the differences. One might argue, for example, that it is more likely for parents to reject their children as boys enter the conflictful adolescent period than it is to reject them when they are small children. By this argument, differences between the interviewed and observed groups would not be due to biased reporting, but would be an outcome of the maturational process. Although this interpretation could fit some of the findings, it would not explain others, such as the tendency, within the observed group, for less role differentiation in the family.

REFERENCES

1. HAVIGHURST, R. J., & DAVIS, A. A comparison of the Chicago and Harvard studies of social class differences in child rearing. *Amer. sociol. Rev.*, 1955, 20, 438-442.
2. MACCOBY, E. E., & MACCOBY, N. The interview: a tool of social science. In G. Lindzey (Ed.), *Handbook of social psychology*. Addison-Wesley, 1954. Pp. 449-487.
3. McCORD, W., & McCORD, J. (with ZOLA, I. K.) *Origins of crime*. Columbia Univer. Press, 1959.
4. McCORD, W. & J. (with J. Guderman) *Origins of alcoholism*. Stanford Univer. Press, 1960.
5. McCORD, W., PORTA, J., & McCORD, J. The familial genesis of psychoses. *Psychiatry*, in press.
6. SEARS, R. R., MACCOBY, E. E., & LEVIN, H. *Patterns of child rearing*. Row, Peterson, 1957.

EFFECTS OF MATERNAL EMPLOYMENT ON THE CHILD[1]

Lois Wladis Hoffman*
University of Michigan

Empirical studies of the effects of maternal employment have long suffered from a paucity of adequate controls. It was the lack of controls, for example, which led to the long-standing belief that maternal employment was strongly associated with juvenile delinquency. In fact, both maternal employment and delinquency were associated with social class and with broken homes; and, when the latter variables were controlled the relationship disappeared, suggesting it had been a spurious one (3).

Yet, the new studies on maternal employment, despite their more adequate controls, have often produced more confusion than clarification. Maternal employment has been associated with *less* delinquency, *more* delinquency, withdrawal behavior, dominance behavior, and nothing at all (1, 4, 9, 10, 11). Furthermore, there are existing theories to account for each of these findings. Thus, delinquent and hostile-assertive behavior might result from parental neglect and lack of sufficient discipline due to the mother's absence from the home; from hostility on the part of the child at being deserted; or, in conjunction with Henry's notion (5) that father discipline leads to the expression of hostility outward whereas mother discipline leads to intrapunitiveness, from father discipline replacing mother discipline.

Withdrawal and dependency responses by the child might also be expected. Loss of mother through death or hospitalization has been associated

---

* Research Center for Group Dynamics, University of Michigan, Ann Arbor.

[1] This paper was presented at the National Council on Family Relations, Ames, Iowa, 1959. The research was supported by the National Institute of Mental Health.

Child Develpm., 1961, 32, 187-197.

with overdependency—the young child perhaps seeking in this crisis a return of the maternal relationship he had in the past. Something similar might occur as the result of the mother's spending a great deal of time outside the home working. Still another reason for withdrawal and dependency might be that the working mother feels guilty about working outside the home and responds with a pattern of "smother love" or overprotection. Such a pattern could alleviate the mother's guilt. In addition, if it is true that working mothers have greater power needs than nonworking mothers, it would also be an appropriate pattern for expressing power over the child in a way that appears to be benign and admirable.

Some, such as Cyrus (2), would expect working mothers to be less frustrated and more capable of warm interaction. Others would say that the energy output required by trying to fill two jobs would make the mothers more irritable with their children.

Instead of speculating as to which of these theories is correct, the present paper will suggest and illustrate an approach in which the question is: *When does one pattern operate and when does another?*

Each theory assumes something about the working situation that is sometimes, but not always, correct; and, when certain assumptions hold, different mother-child interactions will result and the child will be affected accordingly. One important factor implicit in several of the theories involves the question of whether or not the mother enjoys her employment. It is this difference that the present study has selected for focus. It was expected that the effects of employment on the mother-child relationship, and consequently on the child, would be different when the working mother enjoyed work than when she did not.

Two of the theories mentioned above, the "guilt-overprotection" theory and the "neglect" theory, lead to very different predictions about the mother's behavior toward the child, and, although both theories predict that the working mother's child will be disturbed, the pattern of disturbance predicted by each is very different. Both these theories served as guides in the present investigation. The "guilt-overprotection" theory was expected to operate for the working women who enjoyed work, and the "neglect" theory for the working women who did not enjoy work.

Only the working woman who enjoys her employment should be guilty about it, and as such it was predicted that such a woman would try to compensate for her employment by showing a great deal of affection toward the child, disciplining the child adequately to leniently, and being careful that the child should not be inconvenienced by the fact that she works, e.g., by having to help with household tasks. The child correspondingly should be relatively nonhostile, nonassertive, and, if this parental pattern were carried to the point of overprotection, somewhat withdrawn and passive. The working woman who dislikes her work should, on the other hand, be quite guiltless. As such, she was expected to show more of a withdrawal from the maternal role—showing less positive affect and less disciplining. She should feel freer to inconvenience the child—specifically to expect the child to help with household tasks. The child, feeling somewhat resentful and lacking sufficient discipline, was expected to be assertive and hostile. The predictions were, then, that both groups would show different patterns than when the mother was not employed, but that these differences would not be the same.

PROCEDURE

The sample included 176 white, *intact* families with at least one child in the third through sixth grades of three elementary schools in Detroit. There were 88 working-mother families and 88 nonworking-mother families matched on occupation of father, sex of child, and ordinal position, including the important status of being an only child. Except for sex of child, all matched variables were selected because they were known to be related to maternal employment (7) and because it was believed they might be related to the dependent variables. Pairs of families were matched by sex of child so that it might be possible to examine the relationship between maternal employment and the dependent variables separately for boys and girls, to be certain that combining was legitimate. All statistical comparisons are between working and nonworking populations, and with a few exceptions which will be pointed out later, *all comparisons are for matched pairs.*

The data reported here are based on questionnaires filled out by the children, interviews with the mothers, teacher ratings, and a classroom sociometric. Each of these measures will be taken up in turn when the findings are presented.

In the interview with the mothers, the working mothers were asked, "How do you feel about working?" These answers were coded as to the predominance of positive or negative attitudes toward work. Responses to this question were on the positive side for most respondents. Sixty-five respondents were classified as liking work, and 23 as disliking it.

All hypotheses were tested separately for the two groups, i.e., by comparing the 65 mothers who liked work with their 65 matched nonworking counterparts and the 23 mothers who disliked work with their 23 matched nonworking counterparts. Significance tests for differences included the *t* test for correlated means where the data were quantitative, sign tests where the data were qualitative, and the *t* test for proportions where the data could not be analyzed by pairs.

RESULTS

Table 1 summarizes the results dealing with the mother's affect and behavior toward the child. The first comparison deals with the child's perception of the mother as a source of positive affect. The child had been asked to indicate the persons who, from his own experience, best fitted each of a series of verbs, such as "praises" and "smiles," and to indicate the

TABLE 1

MOTHER'S AFFECT AND BEHAVIOR TOWARD CHILD: WORKING-NONWORKING COMPARISONS BY ATTITUDE TOWARD WORK

Variable	*Working mothers with positive attitude toward work*	*Working mothers with negative attitude toward work*
Child's report		
positive affect from mother	more than nonworking*	less than nonworking
coerciveness from mother	less than nonworking	less than nonworking*
Mother's report		
severity of discipline†	less than nonworking*	no difference
power assertion by the mother†	less than nonworking	less than nonworking*
mother's feelings of sympathy†	more than nonworking*	no difference
mother's feelings of hostility†	less than nonworking*	less than nonworking

NOTE.—Findings reported in all tables are based on the t test for significance of difference between correlated means except where indicated.

* Significant at the .05 level, one-tailed test.

† Comparisons not possible between matched pairs. Each working group therefore compared to pooled sample of nonworking mothers and the t test for significance of difference between proportions used.

degree of fitness of person to verb. The perception of positive affect in the mother-child relationship was measured by the extent to which the child associated "mother" with the following words: "praises," "smiles," "listens," "helps," "explains." As can be seen in Table 1, where the working mother indicates a positive attitude toward her work, the child associates more positive affect with the mother than do children in the matched group of nonworking women. This difference is significant. When we compare the children of working mothers who have negative attitudes toward their work to the nonworking group, we find the opposite relation although it is not a significant one.[2]

The second set of findings deals with the child's perception of coerciveness on the part of the mother. This variable was similarly measured using the words, "threatens," "punishes," and "hits." The table shows that the working mother is less likely than the nonworking mother to be associated with coerciveness in both groups and that the relation is significant for the working mothers who dislike work.

The remaining findings in this table are based on the mother's report. In the interview, mothers were asked to give detailed accounts of two recent occasions when they wanted the child to do something that the child did not want to do. These descriptive accounts were coded for degree of severity of discipline, degree of power assertion over the child, sympathy, and hostility. Unfortunately, these descriptions were not available on all of the matched nonworking respondents. To have a large enough *N*, it was therefore necessary to pool the nonworking respondents and compare the two groups of working women separately with this pooled group. These are not, then, matched comparisons, although the working and nonworking groups are, on the whole, comparable with respect to the matching variables. The findings show that where the mother has positive attitudes toward employment she is less severe in her discipline, uses less power-assertive influence techniques, and feels more sympathy and less hostility during the interaction than is the case with the nonworking group. For the working mothers with negative attitudes toward employment there is less power assertion and somewhat less hostility.

---

[2] The direction of the relation is reported even where differences are small; so, except for the significant relations, differences can be due to chance variations.

Taking the findings in Table 1 as a whole, they tend to support the hypothesis that the working woman who gains pleasure from her work shows the child more affection and uses milder discipline. Although she is emotionally involved in the discipline situation, the involvement is to a great extent one of sympathy for the child. The working mother who does not enjoy her work, and whom we would expect to be relatively guilt-free thereby, also shows a tendency toward mild discipline. However, here it is not part of a larger pattern which includes affection and sympathy toward the child. Instead it seems to be part of a general pattern of less involvement with the child. This suggests that the dislike-work group tends to withdraw from the maternal role.

Table 2 deals with the child's participation in tasks. This was measured by the child's responses to a form of the "Doing Things" questionnaire developed by Herbst (6), in which the child is asked to report the extent of his participation in several household tasks (20 tasks in this case). As expected, the children of the working mothers who do not enjoy their work participate more in household tasks than do the children of the nonworking mothers; the children of the working mothers who enjoy work participate

TABLE 2

CHILD'S PARTICIPATION IN HOUSEHOLD TASKS: WORKING-NONWORKING COMPARISONS BY ATTITUDE TOWARD WORK

Variable	*Working mothers with positive attitude toward work*	*Working mothers with negative attitude toward work*
Child's report		
regular participation in household tasks	less than nonworking	more than nonworking*

* Significant at the .05 level, one-tailed test.

TABLE 3

CHILD'S BEHAVIOR: WORKING-NONWORKING COMPARISONS BY ATTITUDE TOWARD WORK

Variable	*Working mothers with positive attitude toward work*	*Working mothers with negative attitude toward work*
Mother's report		
child's assertiveness toward mother[‡]	no difference	more than nonworking[*]
child's playing with younger children rather than own age or older[§]	more than nonworking[*]	no difference
Teacher's report		
initiation of friendships	less than nonworking[*]	less than nonworking
influence attempts made to peers	less than nonworking	more than nonworking
influence success with peers	less than nonworking[*]	no difference
teacher dependency (girls)	no difference	no difference
teacher dependency (boys)	more than nonworking	more than nonworking[†]
aggressiveness	no difference	more than nonworking
use of physical force	less than nonworking	more than nonworking[*]
impulse control	no difference	less than nonworking[*]

[*] Significant at the .05 level, one-tailed test.

[†] Significant at the .05 level, two-tailed test, difference not predicted.

[‡] Comparisons not possible between matched pairs. Each working group therefore compared to pooled sample of nonworking mothers and the t test for significance of difference between proportions used.

[§] Sign test used for this variable.

less. This is in keeping with the notion that the mothers who enjoy work go out of their way to avoid inconveniencing their children because of their employment.

Table 3 deals with the child's behavior. The child's assertiveness toward the mother is based on the mother's account of the conflict situations in which she wanted the child to do something that the child did not want to do. The child's behavior was considered assertive when he directly indicated that he did not wish to comply and that he did not expect to do so. These behaviors included requests such as "Is it all right to stay up until the end of this program?"—as well as statements of intention such as "No. I am not going to bed until the end of this program."—where the context made it clear that the child was not "just talking." Pleading, crying, complying with complaints or silent anger, and devious attempts to obtain ends ("I have a stomach ache.") were all considered less assertive. The child's tendency to play with younger children is based on the mother's report that his playmates were "mostly older," "mostly younger," or "the same age."

The remaining variables reported in Table 3 are based on teacher ratings. These ratings are described more fully elsewhere (8).

It was hypothesized that the children of working mothers who have positive attitudes toward employment would be nonhostile, nonassertive, and possibly withdrawn and passive as compared with their nonworking counterparts, but that the children of working women who do not like work would be assertive and hostile. The data seem to bear this out. The children whose mothers like work differ in their general lack of initiative-taking and tend to play with children younger than themselves. The children whose mothers do not like work, on the other hand, show assertive behavior toward their mothers and toward their peers. This assertiveness toward peers includes aggressiveness, use of physical force, and less impulse control in general. Only teacher dependency does not operate as expected, the boys in both groups showing more dependency, significantly more in the dislike-work group.

Both the "guilt-overprotection" theory and the "neglect" theory lead to the prediction that the children of working mothers will be more disturbed in general than the children of nonworking mothers. This was tested by comparing the matched pairs on the use of nonadaptive responses to frustration, intellectual performance, and sociometric ratings by classmates. The results of these tests are reported in Table 4. The first measure was based on a teacher rating of how the child typically reacts to frustration. "Realistic acceptance of insolvable problem" and "attacks problem directly" were considered adaptive responses, while responses such as "cries," "goes to pieces," "blames others," and "becomes overcritical of self" were considered nonadaptive. Both groups of working-mother children were more likely to use nonadaptive responses to frustration. They also showed lower intel-

TABLE 4

CHILD'S ADJUSTMENT:
WORKING-NONWORKING COMPARISONS BY ATTITUDE TOWARD WORK

Variable	*Working mothers with positive attitude toward work*	*Working mothers with negative attitude toward work*
Teacher's report		
nonadaptive response to frustration†	more than nonworking	more than nonworking*
intellectual performance	less than nonworking*	less than nonworking*
Classroom sociometric		
liking by others	less than nonworking	less than nonworking

* Significant at the .05 level, one-tailed test.
† Sign test used for this variable.

lectual performance and were somewhat less liked by the other children in the class.

The lower intellectual performance of children whose mothers work is a particularly interesting finding. Being low on performance may be a function of low ability or low motivation. If it is true that the working mothers who like work are guilty about their employment and consequently overprotect their children, then the intellectual *ability* of their children may be impaired. Mothers who solve their children's problems for them may hamper their intellectual development by depriving them of valuable problem-solving experience. Thus, these children would be expected to be low on ability as well as performance. The low performance of the dislike-work children, on the other hand, might be part of the rebellious pattern that seems to characterize this group; here we would not necessarily expect low ability. Scores on the Primary Learning Aptitude Test for third graders and the Detroit Alpha Intelligence Test for fourth, fifth, and sixth graders were available. The expectations were borne out. Only the children whose mothers were positive about their work obtained lower scores on these tests than their nonworking counterparts, and this difference was significant. The others showed no such difference.

SUMMARY AND CONCLUSIONS

The over-all pattern of findings suggests that the working mother who likes working is relatively high on positive affect toward the child, uses mild discipline, and tends to avoid inconveniencing the child with household tasks; the child is relative nonassertive and ineffective. The working mother who dislikes working, on the other hand, seems less involved with the child altogether and obtains the child's help with tasks; the child is assertive and hostile.

The data do not show conclusively that the reason for these different patterns is the presence or absence of guilt in the mother, but they are consistent with such an interpretation. If one were to argue that the causality is reversed, that the attitude toward work is a function of the effect of employment on the child or that happier women produce happier children, it would not be consistent with the fact that both groups of working-mother children show indications of maladjustment. It is only that they show different *syndromes* of maladjustment. Furthermore, in Table 5 it can be seen that the attitude toward employment is in large part a function of the particular job the woman holds.[3]

It is interesting to note in Table 5 that, while attitude toward employment is not purely a function of social class, there is somewhat of a tendency for higher status jobs to be well liked. Because middle-class women,

---

[3] The data presented in Table 5 are consistent with the data reported by Weiss and Samelson (12). Their data also show, for example, a positive attitude toward employment on the part of domestic workers.

TABLE 5

OCCUPATION AND ATTITUDE TOWARD WORK

Occupation	*Working mothers with positive attitude toward work*	*Working mothers with negative attitude toward work*
Professional and semiprofessional, e.g., teacher, social worker, registered nurse	14	1
Skilled, e.g., secretary, bookkeeper	11	7
Sales	14	0
Waitress	1	4
Semiskilled or unskilled factory and clerical	20	10
Domestic	5	1
Total	65	23

compared to lower-class, can obtain higher status jobs, because their employment is likely to be choice rather than necessity, and because there is probably greater anxiey about childrearing in the middle class, it is likely that guilt and overprotection are more a middle-class response to maternal employment while neglect is more a lower-class response.

The real purpose of this paper, however, is not so much to elaborate on any particular theory as to suggest an approach to the study of the effects of maternal employment. The need to control on spurious variables has been established. The problem now is to differentiate the all-inclusive variable of maternal employment into more highly specified categories. Only in this way can we come to understand the processes by means of which employment affects the child. If one wants to know whether or not the sheer absence of the mother from the home is the important variable, then one should subdivide the working women according to the number of hours they are away from home while the child is there. If one wants to know the importance of the adequacy of supervision, then one should subdivide the women according to whether or not the substitute supervision provided is adequate. One might feel that the important variable is the father's reaction to the mother's employment. In this case, the working women can be subdivided according to some measure of the father's attitudinal or behavioral response to his wife's employment. In short, by selecting a particular aspect of mother's employment, differentiating the working mothers according to this variable, and examining the differences between working and nonworking groups separately, one can make a more formidable step toward understanding the crucial aspects of maternal employment in relation to the child.

There are still other advantages of this kind of approach. When the working-mother group is not differentiated, it includes a heterogeneous group and, as such, the working-nonworking comparisons will obscure differences due to counteracting subgroup relations. For example, in the data reported here, since one group of working-mother children was less assertive and the other group was more assertive, the two tendencies canceled one another out. Thus, when the total group of working-mother children was compared with the matched nonworking group, there appeared to be no differences in child assertiveness. The same is true for the participation of children in household tasks.

Another kind of error that can result from failure to differentiate the working mothers is when what seems to be a single pattern of findings actually includes two or more distinct patterns. For example, when the total groups were compared, working mothers' children were found to play more with younger children and also to show less impulse control. Doing separate analyses for the two groups of working mothers, however, revealed that these results did not come from the same subjects. Although it was not true of the present data to any great extent, it would be possible for a pattern to emerge from the general working-nonworking comparisons which actually characterized neither the total group nor any single subgroup, but was entirely a coincidence of findings contributed by different clusters of subjects.

One final point should be made. In most working-mother studies, relations are examined between maternal employment and some aspect of the child's behavior *directly,* without considering intervening variables such as the effect of maternal employment on parent-child relations, family structure, or other aspects of family life. The jump between maternal employment and child behavior is too broad to be covered in one leap. One must first understand the various steps that intervene.

In summary, this paper suggests that maternal employment has a different effect on the mother-child relationship and on the child's behavior depending on whether or not the mother enjoys working. The purpose in doing this analysis was to show that greater understanding of the effects of maternal employment will come about when working mothers are differentiated along some theoretically relevant dimension and when the effects on family life, as well as on the child, are examined.

REFERENCES

1. BANDURA, A., & WALTERS, R. H. *Adolescent aggression.* Ronald Press, 1959.
2. CYRUS, D. Problems of the modern homemaker-mother. In J. T. Landis and M. G. Landis (Eds.) *Readings in marriage and the family.* Prentice-Hall, 1952. Pp. 392-402.
3. GLUECK, S., & GLUECK, E. Working mothers and delinquency. *Ment. Hyg.,* 1957, 327-352.
4. HAND, H. B. Working mothers and maladjusted children. *J. educ. Sociol.,* 1957, 30, 245-246.

5. Henry, A. F. Sibling structure and perception of the disciplinary roles of parents. *Sociometry*, 1957, 20, 67-74.

6. Herbst, P. G. Analysis and measurement of a situation: The child in the family. *Hum. Rel.*, 1953, 6, 113-140.

7. Hoffman, L. W. Effects of the employment of mothers on parental power relations and the division of household tasks. *Marriage Fam. Living*, 1960, 22, 27-35.

8. Hoffman, L. W., Rosen, S., & Lippitt, R. Parental coerciveness, child autonomy and child's role at school. *Sociometry*, 1960, 23, 15-22.

9. Nye, F. *Family relationships and delinquent behavior.* Wiley, 1958.

10. Rouman, J. School children's problems as related to parental factors. *J. educ. Res.*, 1956, 50, 105-112.

11. Siegel, A. E., Stolz, L. M., Hitchcock, E. A., & Adamson, J. Dependence and independence in the children of working mothers. *Child Develpm.*, 1959, 30, 533-547.

12. Weiss, R. S., & Samelson, N. M. Social roles of American women: their contribution to a sense of usefulness and importance. *Marriage Fam. Living*, 1958, 20, 358-366.

TEST ANXIETY AND CLASSROOM OBSERVATIONS[1]

KENNETH S. DAVIDSON* *and* SEYMOUR B. SARASON
Yale University

In a previous publication (4) we have described a variety of studies comparing high anxious (HA) and low anxious (LA) children. The measure primarily used for the assessment of anxiety level was the Test Anxiety Scale for Children (TASC). The pattern of these previous findings has been clearly and extensively less favorable for the high anxious children, particularly for the high anxious boys.

In one study (3) classroom observations were made of HA and LA children matched for grade, sex, and intelligence. Those observations described the HA children (particularly the boys) as less secure, task-oriented, and academically adequate than the LA subjects. Because those classroom observations were primarily exploratory and one of a series of procedures for studying 32 matched pairs of HA and LA children, each child was observed for only one hour. The data obtained from these observations were necessarily limited in scope and could not clarify the nature of dynamics and changes occurring in children that would aid in understanding the process by which anxiety interferes with scholastic effectiveness. In that study the process of subject selection did not permit assessment of the effects resulting from differences in teachers and classroom atmospheres.

As a result of these considerations the classroom observations in the present studies had two purposes: (a) to determine the relation of anxiety about school to a wide variety of personality and behavior variables measured in the classroom and (b) to explore the effects of differences in classroom atmosphere upon those relations.

PROCEDURE

The present study was carried out in the three second grade classrooms of a North Haven, Connecticut, elementary school. The choice of the second grade also permitted study of the important question of how early high

---

* Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut.

[1] The research reported herein was performed pursuant to a contract with the United States Office of Education and a grant from the National Institute of Mental Health, Department of Health, Education, and Welfare.

Child Develpm., 1961, 32, 199-210.

anxious children can be detected in order to initiate early steps towards prevention of negative effects of anxiety on learning and development. The TASC can be administered in the usual group procedure to most children at the end of first grade. In the last month of the previous school year the TASC and a similarly constructed and administered questionnaire, the Defensiveness Scale for Children (DSC),[2] were given to the first grade children.

The DSC includes the 11 Lie items formerly found in the General Anxiety Scale for Children (GASC) (4) and 27 other items. Children who give "Yes" answers to questions of the DSC are saying in effect that they can admit and communicate to others about a variety of reactions such as anxiety, shame, aggression, insecurity, and concern about themselves and their relations with others. Children who answer "No" more frequently would generally be more "defensive." The interest in a "defensiveness" score was an exploratory one and stemmed from various considerations. For example HA children were able to admit to feelings of anxiety on the TASC and GASC, but parental interview findings (1, 4) indicated that they did not communicate freely about their feelings to their parents. Also, some LA children with high Lie scores seemed to communicate in a defensive manner that was more typical of HA children. Finally, the TASC discriminated very effectively between HA and LA boys but significant results were much less frequent for girls. The hope was that the DSC would differentiate among girls more effectively or aid in understanding the limited results obtained for girls with the TASC. For these reasons the defensiveness scores were obtained and correlated, in an exploratory manner, with the same measures of personality and classroom behaviors as were the TASC scores.

All of the children in the three second grade classrooms were observed by two trained observers who made notes spontaneously. They observed all morning or all afternoon daily for four months beginning the first day of school. In the original plan one observer sat in one room, the other in a second room. Both observed the children in the third room in order to provide a measure of the adequacy of inter-observer agreement in rating children on personality and classroom behavior variables. After a few observation sessions in this third room, both independently reported that the children expressed themselves much less freely in the third room than the children in the other rooms. Each felt they were not getting to understand the children in the third room as well, which indicated that, unless more freedom of expression occurred in those children, the inter-observer reliability in evaluations of those children could be inadequate for subsequent analyses.

As a result, beginning about the third week, both observers sat, at different times, in each of the three second grade rooms and took notes on all the children. The number of children in the study varied because families moved, but the maximum number of subjects in any aspect of the study was 96, including 53 boys and 43 girls, of whom 40 boys and 37 girls had received the TASC and the DSC in first grade.

---

[2] The last item (4) in the references contains the TASC and the GASC as well as a chapter on adminstration of these scales. The DSC can be obtained from the authors for research purposes only.

In order to assess the subjects' personality and classroom behavior, a check list of 25 variables was formed with definitions provided for each of the items. After three months both observers rated all 96 children, and each teacher rated the children in her class on a 10-point scale of each of the 25 items. A number of the items were selected because they had yielded significant results when parents of HA and LA subjects had rated their own child on those qualities (2, 4). Other items dealing with daydreaming, level of achievement, speed of learning, memory, and attention to schoolwork were included because of their possible significance for both anxiety and classroom effectiveness. However, predictions were not made for these items because children's behavior in school may vary considerably from their behavior at home and because parents and teachers may differ in their values, expectations, and sanctions regarding children's behavior and personality.

RESULTS

There were two phases in the analysis of the data obtained in this study. In the first phase the observers' and teachers' check-list ratings were analyzed to determine the degree of inter-observer and teacher-observer agreement. Also, the children's TASC and DSC scores were correlated with the check-list ratings given them by the observers and their teacher. In the second phase the observers' classroom notes were analyzed to compare the teachers' methods of handling children's needs and the teachers' expression of emotions and value judgments.

Personality Check List

Inter-observer reliability. The check-list study began with an assessment of the inter-observer reliability. The two observers' ratings of the children were correlated for each of the 25 items of the check list. These correlations were computed for the following six groups: all 96 subjects, the 53 boys, the 43 girls, and each of the three separate classrooms with boys and girls combined. The results of this analysis are presented in Table 1. All subsequent analyses involving check-list ratings utilized only those 18 items with inter-observer reliability levels of at least .40 for boys and girls both separately and combined. Two major findings emerge from this assessment of reliability: First, there is significantly greater agreement between observers in evaluating boys as compared with girls. On 21 of the 25 items, the reliability coefficient is higher for boys than for girls ($p < .01$). Second, the inter-observer agreement for the children in rooms 2 and 3 is similarly greater than for the children in room 1, which was the room originally chosen for the reliability check. It should also be noted that there are uni-

TABLE 1

INTER-OBSERVER RELIABILITIES[*]

Item	*Children* N=96	*Boys* N=53	*Girls* N=43	*Room 1* N=32	*Room 2* N=32	*Room 3* N=32
1. Anxiety[§]	.27	.36	.23[‡]	.20[‡]	.28[‡]	.32[†]
2. Anxiety facilitates[§]	.48	.65	.08[‡]	.47	.68	.31[†]
3. Independent	.48	.51	.41	.29[‡]	.59	.54
4. Hides emotion	.61	.66	.56	.47	.67	.71
5. Difficulty in communicating	.49	.50	.53	.58	.60	.52
6. Submissive	.70	.79	.54	.60	.73	.80
7. Cautiousness	.69	.71	.58	.58	.66	.71
8. Not sensitive	.46	.50	.50	.13[‡]	.22[‡]	.78
9. Relaxed[§]	.23[†]	.33[†]	.12[‡]	.38[†]	.25[‡]	.09[‡]
10. Unambitious	.47	.44	.50	.30[†]	.53	.49
11. Set in ways[§]	.64	.79	.30[†]	.65	.59	.72
12. Not well liked	.56	.77	.42	.39[†]	.64	.67
13. Immature	.61	.63	.63	.57	.67	.58
14. Sociable[§]	.66	.75	.33[†]	.35[†]	.66	.85
15. Does not daydream	.47	.45	.50	.39[†]	.60	.42
16. Inactive	.62	.74	.46	.63	.73	.46
17. Underachievers	.49	.49	.50	.17[‡]	.76	.71
18. Learns quickly	.88	.74	.83	.74	.86	.82
19. Forgets	.78	.78	.75	.65	.83	.83
20. Does not fear failure[§]	.33	.31[†]	.10[‡]	—.09[‡]	.09[‡]	.52
21. Does not pay attention	.72	.87	.57	.65	.56	.81
22. Weak conscience	.73	.78	.55	.66	.70	.85
23. Masculine	.75	.66	.56	.51	.96	.78
24. Optimistic[§]	.46	.54	.39	.16[‡]	.60	.57
25. Not responsible	.78	.84	.62	.78	.74	.84
Mean (25 items)	.60	.65	.48	.44	.65	.67
Mean (18 items)	.65	.68	.57	.48	.70	.71

[*] All correlations are significant (one-tail test) beyond the .01 level except as noted.

[†] $p < .05$.

[‡] Not significant.

[§] These items were not used in later analyses or in computing means in the bottom row.

formly low levels of agreement between observers about children's anxiety, tension, and fear of failure in comparison to their very adequate agreement on variables such as emotional expression, cautiousness, maturity, speed of learning, retention, and responsibility.

Teacher-observer agreement. As part of this phase of the analysis, teacher-observer agreement was assessed by correlating the teachers' ratings of the children with those by each observer for each of the 18 selected items. These correlations were computed for the three separate rooms with the boys and girls combined in each room. The results are presented in Table 2.

TABLE 2

CORRELATIONS BETWEEN OBSERVERS' AND TEACHERS' CHECK-LIST RATINGS[*]

	TEACHER 1 $N=32$		TEACHER 2 $N=32$		TEACHER 3 $N=32$	
	Rater A	*Rater B*	*Rater A*	*Rater B*	*Rater A*	*Rater B*
1. Independent	.13[‡]	—.06[‡]	.55	.80	.02[‡]	.12[‡]
2. Hides emotions	.16[‡]	.58	.30[†]	.47	.45	.45
3. Difficulty in communicating	.67	.60	.58	.73	.32	.60
4. Submissive	.67	.62	.74	.66	.70	.60
5. Cautious	.69	.37[†]	.59	.72	.76	.68
6. Not sensitive	.25[‡]	.36[†]	.50	.35[†]	.53	.59
7. Unambitious	.69	.52	.54	.32[†]	.60	.52
8. Not well liked	.52	.36[†]	.65	.54	.60	.22[‡]
9. Immature	.23[‡]	.34[†]	.61	.66	.78	.65
10. Does not daydream	.14[‡]	.50	.27[‡]	.49	—.04[‡]	.35[†]
11. Inactive	.53	.67	.78	.80	.71	.39[†]
12. Underachievers	.04[‡]	.45	.56	.64	.31[†]	.51
13. Learns quickly	.68	.65	.79	.84	.73	.78
14. Forgets	.44	.55	.86	.82	.67	.70
15. Does not pay attention	.69	.56	.74	.73	.80	.74
16. Weak conscience	.69	.61	.78	.81	.86	.79
17. Masculine	.83	.73	.91	.71	.68	.72
18. Not responsible	.67	.47	.79	.80	.81	.77
Mean	.52	.51	.68	.69	.62	.59

[*] All correlations are significant (one-tail test) beyond the .01 level except as noted.

[†] $p < .05$.

[‡] Not significant.

The observer-teacher correlations are higher for room 2 than for room 1 for 15 of the 18 items in the case of each observer. Comparing rooms 3 and 1 the correlations are larger in room 3 for 13 of the 18 items in the case of observer A and larger for 10 of the 18 items for observer B. The means of the 18 correlations for each teacher-observer combination show that, for each observer, agreement with the teachers' ratings was greatest for room 2, smaller for room 3, and least for room 1. The difference in mean teacher-observer agreement between room 2 and room 1 was significant for both observers ($p < .01$). Thus, not only the inter-observer reliability but also the teacher-observer agreement is significantly less for the check-list ratings of the children in room 1.

It will be recalled that both observers remarked very early in their observations that the children in room 1 seemed to express themselves less freely than the other boys and girls. This suggested that the observers' and teachers' ratings might be more constricted for the children in room 1 as compared with rooms 2 and 3. Thus, for each item, the standard deviations were computed for each observer's and the teacher's ratings of the children in the separate rooms. Fourteen of the standard deviations for observer A are larger than the corresponding sigmas obtained from the teacher's ratings in room 1. For observer B, 13 of the 18 standard deviations are larger than those obtained for this teacher. However, the outstanding finding is obtained by comparing the standard deviations of the ratings by the three teachers. For all 18 items, the teacher in room 1 constricted her ratings more than the other two teachers and often did so to a marked degree. While the observers tended—observer A only slightly—to constrict their ratings of the children in room 1 more than the other children in the other rooms, the teacher in room 1 constricted her ratings of her pupils more than either observer and much more than the other two teachers.

Relationship of ratings to test anxiety and defensiveness. The last phase of the check-list analysis concerned the relation between the observers' and

TABLE 3

SIGNIFICANT MEAN CORRELATIONS BETWEEN SUBJECTS' TASC SCORES AND RATINGS BY OBSERVERS AND TEACHERS

Variable[*]	*Room 1* $N=27$	*Room 2* $N=24$	*Room 3* $N=26$	*Boys* $N=40$	*Girls* $N=37$	*Children* $N=77$
1. Independent	..	..	..	+.24[‡]	−.26[‡]	..
2. Hides emotions	..	+.25[†]	..	+.27[‡]	−.19[†]	..
3. Difficulty in communicating	..	..	..	+.20[†]	..	..
4. Submissive	..	+.54[‡]	..	+.23[†]	..	+.15[†]
5. Cautious	..	+.47[‡]	..	+.30[‡]	..	+.14[†]
6. Not sensitive	..	−.30[†]	..	..	−.28[‡]	−.16[†]
7. Unambitious	..	..	..	−.20[†]	..	..
8. Not well liked	..	..	..	..	..	..
9. Immature	..	..	..	..	+.22[†]	..
10. Does not daydream	..	..	..	..	..	+.14[†]
11. Inactive	..	+.34[‡]	..	+.25[‡]	..	..
12. Underachievers	..	−.26[†]	..	−.25[‡]	..	−.17[†]
13. Learns quickly	..	..	+.31[‡]	..	..	..
14. Forgets	..	..	+.30[†]	..	..	..
15. Does not pay attention	..	−.47[‡]	..	−.26[‡]	..	−.18[‡]
16. Weak conscience	..	−.58[‡]	..	−.26[‡]	..	−.21[‡]
17. Masculine	..	−.45[‡]	+.24[†]	..	..	−.16[†]
18. Not responsible	..	−.39[‡]	..	−.29[‡]	..	−.14[†]

NOTE.—Mean correlations were computed through z transformations and tested for significance using a two-tail test and the formula, $t_z = M_z\sqrt{3(N-3)}$, with $df = 3(N-3)$.

[*] Positive correlations mean that anxiety is directly related to the variables as listed, while for negative correlations anxiety is inversely related to the listed variables.

[†] $p < .05$.

[‡] $p < .01$.

teachers' check-list ratings of the children and the subjects' TASC and DSC scores. For the children in each room and for the boys and girls separately and combined in all rooms, TASC and DSC scores were correlated with each observer's and the teacher's ratings. For each grouping of children, three correlations for each item were obtained between ratings—the two observers' and the teachers'—and the TASC scores. The means of these three correlations which were significant are presented in Table 3. The same procedure was followed for the DSC, and these results are presented in Table 4.

TABLE 4

SIGNIFICANT MEAN CORRELATIONS BETWEEN SUBJECTS' DSC SCORES AND RATINGS BY OBSERVERS AND TEACHERS

Variable*	*Room 1* $N=27$	*Room 2* $N=24$	*Room 3* $N=26$	*Boys* $N=40$	*Girls* $N=37$	*Children* $N=77$
1. Independent	+.37‡	..	−.26†	..	..	..
2. Hides emotions	+.34‡	−.32‡	..	..	..	..
3. Difficulty in communicating	+.36‡	..	..	..	+.28‡	+.16†
4. Submissive	+.40‡	−.31†	..	..	..	..
5. Cautious	+.54‡	−.28†	−.33‡	..	..	..
6. Not sensitive	+.35‡	..	..	+.22†	..	..
7. Unambitious	..	..	+.36‡	..	..	+.17†
8. Not well liked	..	..	..	..	+.27‡	+.14†
9. Immature	−.40‡	..	+.50‡	..	..	..
10. Does not daydream	..	..	..	..	−.23†	..
11. Inactive	+.26†	..	..	..	..	..
12. Underachievers	..	..	+.46‡	..	+.20†	..
13. Learns quickly	..	..	−.27†	..	−.26‡	−.22‡
14. Forgets	..	..	+.43‡	+.19†	+.32‡	+.25‡
15. Does not pay attention	..	..	+.58‡	..	+.20†	+.18‡
16. Weak conscience	−.42‡	..	+.54‡	..	..	+.13†
17. Masculine	..	..	..	..	+.23†	+.22‡
18. Not responsible	−.31‡	..	+.54‡	..	+.23†	+.17†

NOTE.—Mean correlations were computed through z transformations and tested for significance using a two-tail test and the formula, $t_z = M_z\sqrt{3(N-3)}$, with $df = 3(N-3)$.

* Positive correlations mean that defensiveness is directly related to the variables as listed, while for negative correlations defensiveness is inversely related to the listed variables.

† $p < .05$.

‡ $p < .01$.

Aside from the specific relations between anxiety or defensiveness and the characteristics and behaviors measured by the ratings, two findings are clearly shown. First, TASC scores relate significantly to a whole series of personality characteristics for boys, but not for the girls, while the reverse is true for the DSC. Second, in room 2 anxiety correlates significantly with a number of variables but not defensiveness. The reverse is true in both rooms 1 and 3. The first finding indicates that anxiety about school is related to a variety of personality traits and behavior patterns among boys but only to a few such qualities among girls. On the other hand, high defensive (HD) and low defensive (LD) girls differ in many respects while HD and LD boys tend not to differ much. In other words, anxiety level, as measured by the TASC, is an effective predictor of boys' personality and behavior patterns in school, whereas defensiveness, as measured by the DSC, performs in the same manner for girls.

In view of the variation in both inter-observer and teacher-observer agreement on the check-list ratings, one would expect corresponding results in correlating those ratings with both the TASC and DSC scores. However, in room 1, where there was the least agreement about ratings, the DSC yielded significant results for 10 of the 18 items and no significant results for any of the TASC items. In room 2, where there was greatest agreement about ratings, the TASC yielded significant results for 10 of the 18 items, but the DSC yielded significant results for only three items. In other words, in spite of the lower reliability for ratings of children in room 1, those ratings are significantly correlated with defensiveness. These findings suggest that either anxiety or defensiveness may be the meaningful variable for different classrooms.

Classroom Observation Notes

This phase of the study was concerned with the teachers' methods of coping with the children's needs and personality characteristics. The analysis focused on teachers' handling of children's dependence, cautiousness, fear of failure, motivation, failure and success, and behavior problems (shyness, anxiety, aggression). Instances in which the teacher expressed strong emotions or value judgments were also noted.

The observers independently culled examples from their notes for each teacher for each of these categories. For each incident or note, two clinicians independently indicated by favorable (+) or unfavorable (—) ratings whether the teachers' efforts had helped the children by meeting their needs or had interfered with or reduced the children's effectiveness. Some notes were rated as favorable with unfavorable tones (+ —) or as more unfavorable (— +) while others were given ?'s when both qualities were equal. The clinicians agreed completely on 628 or 70 per cent of their 892 ratings, while disagreeing completely on less than 4 per cent of the ratings. On 745 or 84 per cent of their ratings they agreed about the favorable or unfavorable direction of the rating.[3]

With this extensive agreement the two clinicians held conferences to arrive at a single rating for those notes on which they disagreed. Thus, for

---

[3] The authors wish to express their appreciation to Dr. Wayne Owen for his assistance in these detailed evaluations. In addition the reader may wish to refer to the parental interview study (1; 4, pp. 222-225) in which the same general procedure of "favorable-unfavorable" ratings was used.

Table 5

FREQUENCIES OF POSITIVE AND NEGATIVE RATINGS OF OBSERVERS' NOTES

	Teacher 1		*Teacher 2*		*Teacher 3*	
	+	−	+	−	+	−
Dependency	18	72	98	16	37	31
Caution	5	29	40	9	17	16
Fear of failure	5	11	24	4	13	7
Motivation	11	31	24	5	16	19
Failure and success	16	37	40	1	29	13
Shyness, anxiety, aggression	7	10	29	2	12	26
Emotion and value judgment	2	33	28	7	2	14
Total	64	223	283	44	126	126

each note there was one rating on which the judges were in agreement. The frequencies of favorable (+ and +—) and unfavorable (— and —+) ratings for each variable are given for the three teachers in Table 5.

For all seven categories, the teacher in room 2 received more favorable evaluations by far than either of the other two teachers. For the teacher in room 3, one half of the incidents were rated favorably and one half unfavorably, while only one quarter of the ratings for the teacher in room 1 were favorable. This pattern of results is especially clear with respect to the three teachers' handling of the children's dependence, cautiousness or impulsiveness, and failures or successes, as well as the teachers' expression of emotions and value judgments.

Discussion

While these analyses yielded a series of significant results, there are really two general findings of importance for the study of child behavior and personality in the classroom. First, the evidence indicates that the teacher is a prime variable in such studies, and, second, different patterns of results can be expected for boys as compared with girls.

With respect to the first point, differences were found between teachers with respect to levels of reliability and constriction of check-list ratings, correlation of those ratings with TASC and DSC scores, and the evaluations of observers' notes on the teachers' methods and behavior. In the case of room 1 there was low inter-observer and teacher-observer reliability, constriction, especially by the teacher in rating the children, and unfavorable evaluations of the teacher's behavior. In comparison, for rooms 2 and 3, there was greater inter-observer and teacher-observer agreement and less constriction in the rating of those children, as well as more favorable evaluations of the teachers.

It is distinctly possible that in room 1 the teacher's ineffective handling of the dependence, cautiousness or impulsiveness, failures and successes of her pupils, as well as her negatively evaluated expressions of affect and value judgments, were important determinants of the lack of spontaneous expression and behavior of those children. The notes of each observer and even casual or untrained observations describe this teacher as authoritarian, punitive, and rejecting of needs to be dependent or cautious. As a result, these children did conform to their teacher's rules and suppressed expression of their needs and feelings in a variety of ways.

It should be pointed out that the mean TASC scores were nearly identical for the three rooms. However, the mean DSC score for room 1 was higher, though not significantly, than the means for the other two rooms. Also, the mean Lorge-Thorndike IQ was lowest for room 1, though the difference was not significant. There is the possibility that the greater defensiveness and lower intelligence, in combination, accounts for their less spontaneous expression. It is our own opinion, however, that the greater defensiveness in room 1 may be an effect of the tendencies of the teacher to de-emphasize differences among her children (constriction of ratings) and to misperceive or misinterpret their behavior. That is to say, a teacher with such tendencies may not only be responding ineffectively and inappropriately to the needs and behavior of the children but also may be engendering in them greater defensiveness, i.e., the tendency to suppress the expression of personal feeling. We are not dealing here with a type of teacher-child interaction which occurs infrequently in this particular classroom, but one which goes on over long periods of time. It is also likely that, as this pattern of interaction goes on over time, the defensive behavior of the children affects the teacher in a way so as to reinforce her tendency to misperceive and misinterpret the significance of the behavior of her children.

The final difference between teachers showed defensiveness but not test anxiety related to a whole series of check-list items for the children in rooms 1 and 3, while the reverse was true for room 2. It is possible to speculate on the reasons for this pattern, but, in any event, it suggests that teachers who vary in their methods and behavior may stimulate different feelings and needs in their pupils which then determine their classroom behavior and characteristics. In rooms 1 and 3, receiving expressions of rejection or acceptance, respectively, seemed a central issue for the children, while in room 2 the teacher showed acceptance of all children, though not all behavior, and the central issue seemed to be academic achievement and improvement.

In view of the fact that anxiety was the meaningful variable for boys and defensiveness for the girls, it is possible that an important determinant of the similar result found among the teachers is differential reactions and relations to boys and girls by the teachers. Observers frequently commented that the teacher in room 1 tended to favor girls and the room 3 teacher tended to favor boys, while there was no apparent difference for boys and girls in room 2. In any event, when teachers differ markedly, it seems highly likely that measures of different variables will yield different results for the children in their rooms.

The content of the above paragraph is but an instance of the more general finding of differential results for boys and girls. There was greater inter-observer and teacher-observer agreement in rating boys on the check-list items, and the observers and teachers did not constrict their ratings of the boys as much as they did those of the girls. Since both observers and all three teachers were women, it might be expected that, if anything, they would show more agreement and less constriction of their ratings for girls. That the reverse occurred is probably related to the fact that the observers and teachers agree that, in school, boys give more overt and spontaneous expression to a wider range of feelings and impulses than do girls. This finding does suggest that observation of children as a basis for measurement is likely to be more effective for a wider range of variables for boys than for girls.

The main difference between boys and girls, however, was the finding that test anxiety scores serve as predictors of boys' but not girls' school behavior and characteristics, while the pattern was reversed for the defensiveness scores. The fact is that differences between HA and LA have been found much more frequently for boys than for girls, and these results have generally showed the LA boys in a more favorable light. This led to speculation that feeling and admitting to anxiety is an ego-alien and disturbing state of affairs for boys, but may be ego-syntonic, not disturbing, and perhaps helpful in a motivational sense for girls (3). There are some other observations which bear upon this notion. Frequently, in giving the questionnaires, one sees girls nodding delightedly as they answer "Yes" to TASC items. Boys express pleasure in this situation only if they are shaking their heads sideways as they answer "No." Boys who answer these items "Yes" seem quiet and tense and often smile and express relief when they can give a "No" answer. These behaviors clearly fit in with the notion that being test anxious can be ego-alien for boys and ego-syntonic for girls.

There is no clear-cut behavior difference between boys and girls when they are responding to the DSC. However, this scale generally measures freedom to express feelings of anxiety, shame, aggression, and concern about the feelings of others when such expression is appropriate. Responding "Yes" means to the LD—in one sense—that their behavior is appropriate or "good" (i.e., concerned about others, accepting responsibility for oneself, etc.). "No" answers describe one as being able to cover up feelings or unconcerned about being aggressive or others' reactions—in a sense—a picture of a strong, silent, masculine person. Clearly, behaving in a "good" or appropriate manner would be ego-syntonic for both boys and girls. However, being unconcerned about others' feelings, masculine, or aggressive would be ego-syntonic for boys and ego-alien for girls.

While these points remain in the realm of speculation, nevertheless, they do indicate that the TASC may well measure ego-alien feelings for boys and ego-syntonic feelings for girls while the DSC measures ego-syntonic qualities of boys and ego-alien qualities of girls. If boys or girls can accept equally being high or low with respect to a given variable, that characteristic may be relatively meaningless as a determinant of their behavior and personalities. However, being high or low on a trait that is unacceptable is the difference between having a lot or a little of a disturbing quality or feeling. Anxiety, apparently ego-alien for boys and ego-syntonic for girls, is an effective predictor for the former but not the latter. Defensiveness, which may well measure ego-syntonic qualities for boys and ego-alien traits for girls, is an effective predictor of the behavior and personality characteristics of girls.

References

1. Davidson, K. S. Interviews of parents of high anxious and low anxious children. *Child Develpm.*, 1959, 30, 341-351.
2. Davidson, K. S., Sarason, S. B., Lighthall, F. F., Waite, R. R., & Sarnoff, I. Differences between mothers' and fathers' ratings of low and high anxious children. *Child Develpm.*, 1958, 29, 155-160.
3. Sarason, S. B., Davidson, K. S., Lighthall, F. F., & Waite, R. R. Classroom observations of high and low anxious children. *Child Develpm.*, 1958, 29, 287-295.
4. Sarason, S. B., Davidson, K. S., Lighthall, F. F., Waite, R. R., & Ruebush, B. K. *Test anxiety in elementary school children: a report of research.* Wiley, 1960.

IMITATIVE COLOR CHOICES IN KINDERGARTEN CHILDREN

JUDY F. ROSENBLITH*[1]
Brown University

Miller and Dollard (2) distinguished several types of imitative behavior among which is one they call *copying* and one they call *matched-dependent behavior*. Copying refers to instances in which one person learns to model his behavior on that of another, while matched-dependent imitation refers to instances in which (a) the response of the imitator matches that of the leader and (b) the imitator is dependent on the leader for cues to act or to match behavior. Miller and Dollard assert that both copying and matched-dependent imitation are learned in the same way. Accordingly, one might expect both types of imitative behavior to be related to experimental manipulations in the same way.

The present study deals with matched-dependent imitation as a function of the same set of variables (or experimental conditions) that were related to copying behavior in an earlier report (5). The variables are: (a) the sex of the leader, (b) the sex of the leader in relation to that of the child, and (c) the way in which the leader interacts with the child. These variables were chosen for their relevance to both identification and learning theories.[2]

---

* Psychology Department, Brown University, Providence 12, Rhode Island.

[1] This paper is based on a dissertation submitted to Harvard University in partial fulfillment of the requirements for the Ph.D. The author would like to express gratitude to Dr. Wesley Allinsmith for his encouragement and counsel. This research was aided by a grant to the Laboratory of Human Development from the U. S. Public Health Service.

[2] A fuller discussion of historical and theoretical issues which led to this choice of variables is to be found in earlier reports (4, 5).

Child Develpm., 1961, 32, 211-223.

The design employed in the over-all study permits us to examine both copying, discussed previously (5), and matched-dependent imitative behaviors, not only in relation to the same experimental manipulations, but within the same subjects. Copying was assessed by the amount of improvement in performance on the Porteus Maze Test of Intelligence. This task allowed us to categorize the children on the basis of their level of pretest maze performance. Thus, we had a rough way of controlling for intelligence or mental age, which may be expected to affect a child's ability to benefit from a leader (or model) or may be related to differential altertness to cues provided by a model.

Since we assess the occurrence of matched-dependent imitative behavior (specifically, use by the child of the same *color* of pencil as that used by the model) in a situation focused on other behaviors (namely, solving the maze), we maximize the probability that the child will not perceive color matching as an instrumental response. The design furthermore enabled us to avoid the chance that the model would use the child's preferred color of pencil, since this preference had already been determined at the same time as the level of maze performance had been measured. As a matter of fact, the model's color choices were the exact reverse of the preference pattern shown by the child.

In this experiment, unlike those reported by Miller and Dollard (2), Schein (6), and Schwartz (7), we are not concerned with teaching the child to imitate. Here, even more completely than in the copying part of the experiment (5), we are relying on the fact that, as a result of their past experiences in the processes of socialization, kindergarten children will have learned to make imitative responses. Miller and Dollard pointed out that the tendency to match behavior becomes a secondary drive, a drive to imitate. We are counting on the presence of such a drive. The child is not rewarded for imitating the color choice of the model. The experimenter makes no comment on the color choice of either model or child. Thus, the question of color choice is not brought to the child's attention in any way. As a consequence of all these factors, we may be quite confident that, if imitative color matching occurs, it does not stem from rewards operating in this experimental situation, but from (a) a learned drive to imitate or (b) from generalization from previous learnings or (c) from an established identification with adults (the models used in this experiment).

Questions concerning identification theory that were raised previously (5) are particularly relevant here. In general, one would expect that, the more a child is identified with an adult, the more tendency the child would have to imitate responses of the adult, such as the adult's choice of colored pencils. Therefore, we might expect to find differences in the amount of imitation exhibited as a function of the sex of the leader and/or as a function of the leader's sex in relation to the child's sex. The role of "threat" is relevant in this experiment, if it is relevant to strength of identification. Therefore, in half the cases in the interaction between adult and child, the adult withdrew attention from the child before the imitative behaviors were measured.

This paper will examine not only the relation of matched-dependent imitation to the above mentioned variables, but will also address itself to the question of whether both copying (learning by imitation) and matched-dependent (color matching) imitative behaviors follow the same pattern in response to experimental manipulations. In addition, we shall examine the possibility that learning by imitation of the kind that we might designate as task-oriented (or "instrumental" imitation) may be relatively more sensitive to "threat" than matched-dependent imitation (or "role" imitation) which may prove more sensitive to the reward of attention.

METHOD

The *S*s were 80 kindergarten children from two public schools in a suburb of Boston composed of middle and upper-middle class residents.[3] Each child was led individually from the classroom to the experimental room by the experimenter in order to "play a game." When they arrived, *E* said to the child: "Here is a chair for you, and here is one for me (at the opposite sides of the table). Here are *some pencils* for you. This is a paper and pencil game. The first thing I want you to do is take *one of your pencils* and draw. . . ." There were four pencils in front of the child: one red, one blue, and one green, each producing its color, and a yellow one which made ordinary lead marks. Nothing further was said about the pencils unless the child asked a question. If he did, he was again told that the pencils were all there for him and he could use whichever ones he wanted.

The child was then administered the Porteus Maze Test (3); slight modifications in instructions and procedure were made to render them more suitable to the age of the children.[4] On the basis of their performance in this initial session, children were assigned to one of four groups. Group 1 consisted of those who passed only the first two mazes (i.e., the 3- and 4-year mazes). Group 2 consisted of those who passed a total of three or four of the mazes. Group 3 passed five mazes, and group 4 passed six or more of the mazes. We might note the fact that level of initial performance on the Porteus Maze Test of Intelligence should reflect levels of ability or maturity. Children from each group were then assigned to each of the treatments. In order to have larger *N*s for comparisons, groups 1 and 2 will often be considered together in relation to groups 3 and 4.

One to three weeks after this initial session the child was brought from his room by the *E* to have "a second turn at the game." Just before arriving at the experimental room, the *E* said to each child: "You know what? I have someone else here to play the game with us." *E* and child then entered the room, and *E* introduced the child to the other adult by first names. *E* then explained to the child that she (*E*) had something to do for a while and invited the child and the other adult to go to another table in the room and play with the toys on it until *E* was ready to play the game.[5] The child was seated facing the table of toys and the adult at the side of the table next to the child in such a way as to face both the table and the child.

---

[3] These are the *S*s who were in one of the treatment groups that had a leader in the previous study (5).

[4] Description of the exact procedure can be found in (4).

It should be noted that *E* essentially was treating the other adult in the same manner as the child. *E* did nothing to enhance the adultness (or maleness or femaleness) of the leaders except to reward their proficiency in performance. They were not called Mr. or Mrs. and not described as fathers or mothers. No attention was called to their choice of pencil color. Each leader interacted with the child, however, as an adult would and not as another child would.

The different treatment conditions were: I. Male leader, child of the same sex, leader attentive to child throughout the 10-minute period that preceded the copying session proper. II. Male leader, child of the same sex, leader attentive to the child for five minutes, then leader said: "I'm sorry, I can't play with you any more, I have to read a book." Leader then turned his chair so that his back was toward the child and read a book for five minutes. III. Male leader, child of the opposite sex, attentive throughout (as in I). IV. Male leader, child of the opposite sex, withdrew attention after five minutes (as in II). This entire design was then repeated using a female leader.

At the end of the 10-minute period *E* said: "I'm ready to play the game now." Adult leader and child then took adjacent seats at the table on the side opposite *E* and facing her. Both the adult and the child had four colored pencils at their places as the child had had in the previous session. *E* said to the child: "You've played this game with me before. ——(adult) wasn't here when we played then, so how would it be if you take turns and give ——(adult) the first turn." The adult was then given the first maze which the child failed previously and the same instructions as were given the child. The leader picked up a pencil (of the color least used by the child in the initial session) and did the maze correctly at a rather slow pace. The child watched during this period and then was given a turn on the same maze. There were two such trials on each maze which the child had initially failed. After the third trial the adult shifted to the second least used color and continued to shift after every three trials. This was continued until the criterion for discontinuing testing was reached. The proportion of all trials on which the child used the same color of pencil as the adult provides the basic data. These data are shown in Table 1 which also depicts the design of the experiment.

---

[5] The toys were selected to have relevance to the questions of identification. They included more or less matched "male" and "female" items: make-up and shaving sets, carpenter and kitchen tools, doll family, and others.

TABLE 1

PERCENTAGE OF SAME COLOR CHOICES UNDER THE VARIOUS CONDITIONS

Groups	BOYS Attention	BOYS Withdrawal	GIRLS Attention	GIRLS Withdrawal
Male Leader				
1	10	11	50	28
2	35	26	54	20
3	33	60	79	13
4	40	66	14	39
All groups	29	37	50	24
Female Leader				
1	20	7	52	54
2	29	14	50	13
3	37	66	31	31
4	19	29	25	33
All groups	27	26	41	28

RESULTS

Sex of Subjects

We shall report on sex differences according to both sex of subject and sex of leader. For girls, 36 per cent of all color choices were imitative; for boys, 30 per cent ($p < .10$).[6] The differences in amount of imitation shown by boys and girls depended on the initial level of performance in part. Girls showed more color matched choices than boys ($p < .001$) in groups 1 and 2 (the poorer, less bright or able groups on initial level of performance). Boys showed more color matched choices than girls ($p < .07$) when initial level of maze performance placed them in the two top groups (3 and 4). The over-all tendency of the girls to show more color matching than boys occurred both with the male leader and with the female leader, but was not significant with either. This nonsignificant tendency for girls to exhibit more color matching than boys is in contrast to the greater learning-by-imitation of the boys who achieved 101 new passes on the mazes to the girls' 69 (*see* Table 2).

Sex of Leader

Thirty-five per cent of all color choices were the same as those of the male leader; 30 per cent were the same as those of the female leader. This difference is significant at the .20 level only. Here again, initial level of maze

---

[6] Probability levels are based on the critical ratio or z test.

TABLE 2

SEX DIFFERENCES IN PERCENTAGE OF IMITATIVE COLOR CHOICES

Groups	*Boys*	*Girls*	*p**
1 & 2	21	40	.001
3 & 4	43	31	.07
All	30	36	.10

* In this and subsequent tables the probability level is determined using the critical ratio or z test.

performance made a difference. The male leader elicited more color matching than the female leader from the brighter *S*s (p is .07 for *S*s whose initial level of performance put them in groups 3 or 4 and .04 for those in group 4); the brighter the *S*s, the smaller the p value (*see* Table 3).

Level of Initial Performance on the Mazes

The interaction of this variable with the other variables was examined. An analysis of variance by ranks, the expansion of the Kruskal-Wallis test used in (5), showed that only the interaction between levels and sex of subject was significant ($p < .01$).

Attention and Withdrawal of Attention

For girls (with both male and female leaders) attention from the leader resulted in more color-matching behavior than did withdrawal of attention. This was true across all levels of performance ($p < .002$) (*see* Table 4). (*See also* Table 5). For *all* boys (with both male and female leaders) there was no significant tendency for attention or withdrawal to be more effective. This result reflects two opposing tendencies. Boys whose initial level of performance was highest (groups 3 and 4) showed a significant ($p < .02$) tendency for withdrawal of attention to be more effective than attention throughout. On the other hand, boys whose initial level of performance

TABLE 3

PERCENTAGE OF SAME COLOR CHOICES WITH MALE AND FEMALE LEADERS

Groups	*Male Leader*	*Female Leader*	*p*
1 & 2	30	29	*ns*
3 & 4	43	32	.07
4	40	23	.04
All	35	30	.20

Table 4

PERCENTAGE OF IMITATIVE RESPONSES IN RELATION TO ATTENTION AND WITHDRAWAL OF ATTENTION

Groups	A*	WA*	*p*	A*	WA*	*p*
1 & 2	24	16	.17	52	27	.003
3 & 4	32	54	.02	38	25	.16
All	28	32	*ns*	46	26	.002

* In this and subsequent tables, A stands for attentive throughout, and WA for withdraws attention half the time.

placed them in groups 1 and 2 showed a nonsignificant tendency for attention to be more effective.

For all boys with the female leader, there was no effect of attention versus withdrawal of attention. There was a slight tendency ($p = .18$) for boys in groups 3 and 4 combined to do more color matching when the female leader had withdrawn attention. For all boys with the male leader there was a nonsignificant tendency for withdrawal of attention to be more effective. The superiority of withdrawal of attention has, however, a *p* value of less than .05 for boys in groups 3 and 4 (*see* Table 5).

Sex of Leader Relative to Sex of Child

Boys showed more imitation of the leader of the same sex while girls showed more imitation for an attentive leader regardless of sex. Taking boys and girls together, we find that the leader of the opposite sex who was attentive elicited more color matching ($p < .03$) than the leader of the opposite sex who withdrew attention. The leader of the same sex who was attentive did not differ from the leader of the same sex who withdrew atten-

Table 5

PROBABILITIES* THAT DIFFERENCES BETWEEN TREATMENTS COULD BE OBTAINED BY CHANCE

	Male Leader				Female Leader			
	Boys		*Girls*		*Boys*		*Girls*	
Groups	A	WA	A	WA	A	WA	A	WA
1 & 2	*ns*		.01		.11		.10	
3 & 4		.05	.10			.18	*ns*	
All		*ns*	.002		*ns*		.14	

* Values are placed under the treatment that was more effective.

tion, but each of them tended to elicit more color matching ($p = .12$ and .15) than did the opposite-sexed leader who withdrew attention (*see* Table 6).

TABLE 6

EFFECTIVENESS OF TREATMENTS WITH SEX OF LEADER CONSIDERED RELATIVE TO SEX OF CHILD

Conditions Compared		*p*
OS* - A	with OS - WA	.03
SS* - WA	with OS - WA	.15
SS - A	with OS - WA	.12

* In this and subsequent tables, OS stands for leader of opposite sex to child and SS for leader of same sex.

The rank order of effectiveness of treatments for enhancing this color-matching form of imitation is shown in Table 7. Figures permitting comparison with the rank order of effectiveness of treatments for learning by imitation are also given.

TABLE 7

RANK ORDERS OF EFFECTIVENESS OF TREATMENTS FOR COLOR MATCHING AND FOR LEARNING-BY-IMITATION*

	SS - A	SS - WA	OS - A	OS - WA
Boys	2 (3)	1 (1.5)	3.5 (1.5)	3.5 (4)
Girls	2 (3)	3 (4)	1 (1.5)	4 (1.5)

* The rank order for learning by imitation is given by the numbers in parentheses.

Number of Pencils Used

A form of behavior other than color matching can be looked at in connection with the influence of the adult model, namely, how many different pencils did the child use. Of the 80 children, 41 used the same color on all or all but one of the pretest trials. Only 15 children did so in the copying session. Every one of these 15 was a child who had used only one color in the pretest session, but only four of them used the same color they had used in the pretest.[7] Of the 15 who still used only one color, 11 had a leader who withdrew attention. Of the 26 who gave up using only one color, 17 had a leader who had been attentive throughout. Of the nine who gave up using only one color *and* who had a leader who withdrew attention, seven had a male leader. Attention leads to giving up the use of one color only; withdrawal of attention by a male leader may lead to the use of several colors of pencils.

---

[7] The number of children who used each of the colors exclusively in the pretest are: red 16, regular 14, blue 10, and green 1.

A χ^2 test can be used to examine the relation between giving up one color and attentiveness of the leader; χ^2 (corrected for continuity) for the 2 by 2 table is 4.26, which for 1 degree of freedom has a p value of .04. Whether the leader was male or female and whether the leader was of the same or opposite sex were, however, not significant factors.

Discussion

It is interesting to compare these data on color matching with those on learning by imitation (4, 5). Hence, the main task of our discussion will be to place the two sets of results in apposition.

1. Sex of subject: Boys tended to show more learning by imitation and girls tended to show more color matching (matched-dependent or "role" oriented imitation).

2. Sex of leader: Whereas the male leader was more effective in producing learning by imitation, there was only a tendency for the male leader to evoke more same-color pencil choices.

3. Attention vs. withdrawal of attention: For girls there was a greater tendency for both kinds of imitation to occur when *S*s had attention than when the leader withdrew attention. In the case of girls with the male leader, the attentive treatment produced more same color pencil choices than did withdrawal of attention. Attention also produced more learning by imitation. However, in the case of learning by imitation the male leader who withdrew attention from the girl was the second most effective condition (essentially tied with the most effective), whereas for color choices this treatment yielded the smallest percentage of same color choices.

For boys the male leader's withdrawal of attention produced a tendency toward more learning by imitation and more same color choices than did attention throughout. However, the female leader who was attentive was more effective than the female leader who withdrew attention, both for learning by imitation and color matching. The female leader who withdrew attention was the least effective of all treatment conditions for both learning by imitation and color matching.

For both sexes of *S*s attention was more effective than withdrawal of attention in leading the child to give up the use of only one color of pencil.

4. Leader of the same or opposite sex to the child: Boys showed more color matching for a leader of the same sex regardless of treatment, but best performance in learning by imitation was for the same sexed leader who withdrew attention and the opposite sexed leader who was attentive. Girls showed more color matching (as indicated in 3 above) for an attentive leader regardless of sex. On the other hand, they had shown more learning

by imitation for the opposite sexed leader regardless of treatment. Taking boys and girls together, the effectiveness of the treatments was shown in Table 6.

5. Initial level of performance of *Ss*: In the color matching data, levels of initial performance interact significantly with the sex of the *Ss* ($p < .01$). In the learning by imitation aspect of the study the effect of initial level of performance was significant at the .07 level. Initial level of performance then affects behavior in both types of imitation, but the direction of the effects are not the same. One might say that the brighter girls are more like the boys in that they improve more in the learning by imitation. Conversely, one might say that the brighter boys are more like the girls in that they color match more. Actually, bright boys, less bright girls, bright girls, and less bright boys color match in that order, whereas bright boys, less bright boys, bright girls, and less bright girls learn by imitation in that order (*see* Table 8).

TABLE 8

PERFORMANCE IN IMITATION AS A FUNCTION OF TYPE OF IMITATION, SEX OF SUBJECTS, AND INITIAL LEVEL OF PERFORMANCE

	BOYS		GIRLS	
Groups	*Same Color Choices*	*New Passes*	*Same Color Choices*	*New Passes*
	PERCENTAGE	AVERAGE NO.	PERCENTAGE	AVERAGE NO.
1 & 2	21	2.3	40	1.5
3 & 4	43	2.8	30	1.9

At this point we should return to the theoretical outlooks which were discussed in this paper and in "Learning by Imitation in Kindergarten Children" (5) and which were the origins of the variables investigated. We can now examine them in the light of the findings of these studies.

First, the reinforcement view of learning which Miller and Dollard hold led us to the expectation that the female should be the most effective model for imitation. This expectation was based on the assumption that five-year-old children (both boys and girls) have been reinforced for imitating more frequently by a female (the mother and now the teacher). Hence, a female should have more secondary reward properties and so elicit more imitative responses. This was not the case either for learning by imitation or for color matching. The author is aware of the possibility that the mother might reinforce the boy for imitating males rather than for imitating her. This, however, would not serve to explain why the female was not the most effective model for the girls for either type of imitation. Indeed, the only condition in which the female leader was highly effective was in the case of boys toward whom she had been attentive throughout the preliminary period.

Second, a general view of identification theory led to the expectation that identification should be sufficiently advanced by kindergarten age that an adult of the same sex as the child should be the most effective model for eliciting imitative responses. This was not generally the case. It *was* the case for boys in the color matching aspect of the experiment.

Third, the role of threat in leading to identification has been considered by a number of theorists. Their considerations led to the expectation that the mild threat involved in withdrawal of attention should lead to more identification and hence to a greater effectiveness of the model who used it, at least with respect to behaviors demanded or suggested by the model. Threat should be particularly effective when it comes from an adult of the same sex as the child. The contingency view of reinforcement might lead to the same expectation. However, in that case, the sex of the person who withdraws attention (threatens the relation) is of less obvious importance. From identification theory, the prediction is clearest for boys with an adult male leader. However, the findings of Hartup (1) led to the expectation that withdrawal of attention from girls by an adult female should be more effective in evoking learning by imitation.

For girls in the matched-dependent imitation study, withdrawal of attention was effective in the opposite way. There was less imitation found in girls who had a leader who withdrew attention. This was true regardless of the sex of the leader. The only counter tendency was for the male leader who withdrew attention to be effective in evoking learning by imitation.

For boys, both the color matching and learning by imitation experiments showed a tendency for withdrawal of attention from the same sexed leader to lead to more imitation. However, neither difference reached the usual level of statistical significance except for boys who were in the higher (brighter) groups according to initial level of performance on the mazes. With the opposite sexed leader, attention was more effective for learning by imitation. In color matching (which was relatively infrequent) attention or its withdrawal made no difference.

Fourth, consideration of the Miller and Dollard view that all forms of imitation are learned in the same way might lead one to expect both kinds of imitation to relate to the variables in the same way. As we have seen, sometimes they do and sometimes they do not.

Fifth, the author had hypothesized that learning by imitation (task oriented or "instrumental" imitation) might be more sensitive to the withdrawal of attention manipulation (threat), whereas matched dependent imitation ("role" oriented imitation) might be more responsive to attention (reward). The only real support for these notions lies in the data of the girls where attention seemed to be the crucial variable in determining the amount of color matching. There is a kind of counter evidence with boys. For them there was some tendency for threat to have an effect in enhancing imitation, and the effect was the same for both kinds of imitation.

In the light of these findings it appears that the variables we have selected for study do have effects. However, the variables interact, and their effects are not those to be expected from any one of the theoretical approaches considered here. In short, the relations of the variables to the amount of imitation are highly complex, and no existing theory is sufficiently differentiated to account for them. One might summarize loosely by saying that in this situation: For girls, color matching depends on attention from the model. Learning by imitation depends on having a male (opposite sexed) model. For boys, color matching depends on having a male (same sexed) model. Learning by imitation is more complex; it depends on having a same sexed leader who withdraws attention or an opposite sexed leader who is attentive. In all cases it is necessary also to take into account the initial level of performance or ability of the *S*s.

It seems probable that each of these approaches is right in some circumstances and that almost all of them need to be taken into account in formulating a theory adequate to all circumstances. Replication and extension of this type of study appears to be a promising avenue for obtaining data on which a new formulation of the problems of identification and imitation can be based.

SUMMARY

This study was carried out in order to test the hypothesis that matched-dependent imitation would respond to a particular set of experimental manipulations in the same way that learning by imitation or copying did. The experimental manipulations involved the sex of the adult leader to be imitated, the sex of the leader in relation to that of the child, and the way in which the leader treated the child in the 10-minute period preceding the imitation session (attentive throughout vs. withdrawing attention for half of the period). The matched-dependent behavior under study involved color matching. A child could match the color of pencil used by an adult to do the Porteus mazes. There was a pretest session in which the child's color preferences were established in addition to his level of performance on the mazes. There was then an experimental (or copying) session in which the child observed an adult before trying the task. A datum which could be observed by the child (other than the way in which the adult performed on the maze) was the color of pencil which the adult used in doing the maze. In the experimental sessions the actual work on the mazes was preceded by a 10-minute period in which the adult behaved in one of two ways: attentively throughout, or attentively for half the time followed by withdrawal of attention for the remaining half of the time.

The data showed a number of trends in the direction of supporting the above hypothesis, and some contradictory to it. Supplementary evidence showed that attention vs. withdrawal of attention makes a difference in the amount of influence a leader has in changing the child's behavior. There

were a number of interesting differences between the two types of imitation that need to be explored further.

References

1. Hartup, W. W. Nurturance and nurturance-withdrawal in relation to the dependency behavior of preschool children. *Child Develpm.*, 1958, 29, 191-201.
2. Miller, N. E., & Dollard, J. *Social learning and imitation.* Yale Univer. Press, 1941.
3. Porteus, S. D. *The Porteus Maze Test and intelligence.* Pacific Books, 1950.
4. Rosenblith, J. F. Imitation in kindergarten children. Unpublished doctoral dissertation, Radcliffe College, 1958.
5. Rosenblith, J. F. Learning by imitation in kindergarten children. *Child Develpm.*, 1959, 30, 69-80.
6. Schein, E. H. The effect of reward on adult imitative behavior. *J. abnorm. soc. Psychol.*, 1954, 49, 389-395.
7. Schwartz, N. An experimental study of imitation: The effects of reward and age. Unpublished senior honors thesis, Radcliffe College, 1953.

NONCONTINUOUS MOTHERING IN INFANCY AND DEVELOPMENT IN LATER CHILDHOOD[1]

D. Bruce Gardner,[*] Glenn R. Hawkes,
and Lee G. Burchinal
Iowa State University

On the strength of assertions which have been made about the effects of inadequate mothering, it might be hypothesized that, if maternal deprivation does occur, the children who experience it should reveal the results of that inadequacy in discriminable variations in personality structure and content. In the absence of such observable, discriminable variations in personality, any assumption of maternal deprivation seems unwarranted, regardless of the fact that unusual mothering conditions may actually have occurred.

In an earlier report (5) a research design was described, the central objective of which was to ascertain both short-term and long-range effects on children of an unusual, noncontinuous mothering experience during infancy. The major condition of that unusual mothering experience was residence in a home management house on a university campus, prior to being placed in private adoptive homes. During the period of residence in the home management house, each infant's care was provided by many different mother figures, each having responsibility for the baby for a few days at a time.

At the same time as the longitudinal study of these subjects has been going on, efforts have been made to locate children who resided in the home management houses over the past few years and to evaluate their present status in comparison with children who did not experience such an unusual mothering situation in infancy. One purpose for including this group of subjects in the over-all study is to make more meaningful any findings

---

[*] Department of Child Development, Iowa State University, Ames.

[1] Journal Paper No. J-3993 of the Iowa Agricultural and Home Economics Experiment Station, Ames, Iowa, Project No. 1299. Acknowledgement is due the Elizabeth McCormick Memorial Fund, for partial financial support, and the Iowa Children's Home Society, for their cooperation. This investigation was supported, in part, by a PHS research grant, M-2963, from the National Institute of Mental Health of the U.S. Public Health Service.

Child Develpm., 1961, 32, 225-234.

obtained with the younger, longitudinal group. In general, the condition of older children who were exposed, as infants, to such unusual mothering should help us to formulate clearer and more significant hypotheses about the future development of the larger group under investigation.

The purpose of this report, then, is to present information on a smaller group of older children who lived for a period of time during infancy in a home management house and who were subsequently adopted into private homes.

PROCEDURE

The subjects for investigation were selected through the cooperation of the Iowa Children's Home Society. That agency was responsible for the adoption proceedings for each home management house baby at Iowa State University since 1940. Their records indicated that there were 62 children who had lived in a home management house and had subsequently been adopted into families in Iowa.

Of that number, however, only 29 were still residing in Iowa and available for the investigation. Their ages ranged from 8 to 17, and their school grade placement from third through twelfth. Data summarizing the experiences of these children prior to being adopted are presented in Table 1.

TABLE 1

PREADOPTION DATA ON HOME MANAGEMENT HOUSE SUBJECTS

	Mean	*SD*	*Range*
Age on admission to H. M. H.*	5.0	3.1	0.5–14.0
Length of residence at H. M. H.	5.1	3.1	0.8–12.0
Age at adoption†	12.1	5.6	3.0–23.0
Number of changes in residence	3.4	1.4	1.0– 7.0

* Age data reported in months.
† Refers to age at which child was first placed in adoptive family.

A group of comparison subjects were selected on a matched-pair basis. That is, a child of the same age and sex who had lived with his own biological family from birth was matched with each adopted child.[2] The matching procedure was carried out in the following manner:

All children in the school class in which the group A child was located were given the Otis Mental Ability Test and were asked to complete a family data sheet. This data sheet provided information about parents' ages, education, occupation, and number of children in the family. In addition, the children completed items from the Gough Social Status Inventory (3), an index of socioeconomic status.

---

[2] For convenience in the discussion which follows, children who had lived in a home management house and, subsequently, in an adoptive family, are designated group A; those who had lived only in their own biological families are referred to as group B.

From the classmates of each group A child, between one and three potential group B children were selected. Factors considered in this selection were: sex, age, Otis IQ, family socioeconomic status, age of parents, parents' educational level, and father's occupation. The families of all children were intact and living together. Because of the selection procedure, each group B child was from the same community, grade level, and class as his mate in group A.

Inasmuch as no one "B" child could represent a perfect match with respect to all of the criteria, the final selection was made by a panel of three judges, working together, who agreed on the one "B" subject who most closely approximated the "A" child when all factors were considered.

TABLE 2

PARTIAL RESULTS OF MATCHING PROCEDURE

Matching Variable	*N*	*Z*	*p*
Otis IQ	29	1.93	.052
Age of children	27	1.22	.222
Age of father	25	3.49	$<.001$
Age of mother	25	4.07	$<.001$
Economic status	29	2.62	$<.01$

Table 2 reveals, in part, the extent to which this matching procedure was successful. Variables for which continuous data were available are listed in the left-hand column. The Wilcoxon matched-pairs signed-ranks test (7) was used to determine whether significant differences remained between the pairs of children. In this nonparametric technique, *N* refers to the number of pairs in which differences occurred. The *Z* score is based on standard deviation scoring units. Probabilities of differences as great as those observed between pairs are shown in the last column.

With respect to sex, the matching was perfect; hence, no test was made. With respect to age, the differences were insignificant. With respect to age of fathers and age of mothers, there were highly significant differences between pairs. Although average scores for groups are not pertinent to the Wilcoxon procedure, they are of interest in interpreting the differences found between pairs. The mean age of fathers in group A (48.4) was considerably higher than that for fathers in group B (41.1). Similarly for mothers, the mean age in group A (45.3) exceeded that for group B (37.7).

Differences between pairs with respect to socioeconomic status favored the adopted children. The absolute differences were small, but consistent enough to be highly reliable.

In spite of the careful efforts to match the pairs on the basis of intelligence, the data in Table 2 indicate that there may be some question about their comparability. The p value of .052 approaches the 5 per cent level, which had been arbitrarily established in advance as the criterion for rejecting the null hypothesis. The mean scores favor the group B children (108.3) over group A (105.6). The differences are certainly not large, and it seems unlikely that such small differences in performance on the Otis examination would be of great significance. We are evidently dealing with the same general range of intellectual functioning for both groups. Thus, it was concluded that the matching procedure had been at least relatively successful with respect to intelligence.

The remaining matching variables included number of children in the family, educational level of fathers and mothers, and father's occupational classification. The occupational classification was based on the six major categories described by North and Hatt (4): (a) professional, (b) proprietary, (c) clerical, (d) craftsmen, (e) operatives, and (f) farmers. Of the 29 pairs, 13 were in agreement with respect to occupational groups. For eight additional pairs, the occupational group of the "B" father was in the next higher or lower classification, adjacent to that of the "A" father. In the remaining eight pairs the two fathers were separated by one or more occupational categories. All of the six occupational categories were represented in both groups, and there was no tendency for one group to rank higher than the other.

Comparison of educational level of parents was accomplished by establishing three broad educational categories: (a) completion of grade school only, (b) completion of high school, and (c) completion of college. Of the 29 pairs, there were 16 agreements and 13 disagreements for fathers, and there were 13 agreements and 16 disagreements for mothers. However, when the subjects were considered as groups rather than as pairs, the proportions in each of the three categories were almost identical, both for fathers and for mothers.

With respect to number of children in the family, the difficulty of maintaining absolute control in matching became even more apparent. Since the "A" children were all adopted and since it is unusual for parents to adopt more than one or two children, it might be expected that important discrepancies would be found here. Nevertheless, of the 29 pairs, 16 were in agreement on the number of children in the family. Another 10 pairs had only one more child in the "B" than in the "A" family, and in just three pairs was there a difference of two in the number of children.

To summarize the results of the matching procedure, the pairs of subjects were perfectly matched with respect to age, sex, educational level, community, and criterion of living in a complete (unbroken) family. Matching on the basis of intelligence was at least relatively successful; although differences between pairs approaching a level of significance were observed, in favor of group B, the average difference between pairs was so small as to warrant the conclusion that we were dealing with the same range of intellectual functioning. Matching of the pairs was less perfect with respect to the factors of size of family, educational level of parents, socioeconomic status of family, father's occupation, and age of parents.

In addition to the forms and tests described above, the subjects completed a number of procedures designed to get at some aspects of personality and adjustment. These tests included the California Test of Personality, the Children's Form of the Manifest Anxiety Scale, and the Rosenzweig Picture-Frustration Study. In addition, data were available for 18 of the pairs for the Iowa Every-Pupil Test of Basic Skills, a widely used test of general school achievement.

With the exception of the Rosenzweig P-F Study, which was administered individually, the tests were given in the classroom group, with all children participating. Thus, the subjects were not aware of the special nature of our interests. All of the tests, excepting the achievement tests, were administered by one of the authors.

Statistical comparisons of the 29 pairs were made with respect to the variables measured by these tests, and the results are presented in the following section.

RESULTS

Subjects were compared in the three major scoring categories of the California Test of Personality: Personal Adjustment, Social Adjustment, and Total Adjustment, the latter representing a favorable balance of elements of the other two. Results, shown in Table 3, indicate that there were no significant differences between pairs in any of these scores. These figures are based on the Wilcoxon procedure, previously described. It will be noted, however, that in the case of the Personal Adjustment scores, the Z value

TABLE 3

SIGNIFICANCE OF DIFFERENCES BETWEEN PAIRS: DEPENDENT VARIABLES

Variable	*N*	*Z*	*p*
California Test of Personality			
Personal Adjustment	28	1.81	.07
Social Adjustment	28	1.39	.16
Total Adjustment	27	1.55	.12
Children's Form, Manifest Anxiety Scale			
Anxiety Score	28	.06	.95
"L" Score	28	.41	.68
Iowa Every-Pupil Test of Basic Skills			
Composite Score	18	.20	.84

of 1.81 and corresponding *p* value of .07 approach the 5 per cent probability level.

The differences favor group B, whose mean score was 60, in comparison with group A, whose mean score was 56. While the statistics do not allow us to reject the null hypothesis, it should be remembered that criteria for rejection, in this case, the .05 level, are always arbitrarily established. Thus, on the basis of this one finding alone, the possibility of there being real differences in personal adjustment between the two groups should not be overlooked.

For the Children's Form of the Manifest Anxiety Scale, no differences attributable to experiences of the group A subjects were observed. This scale provides an index of the level of anxiety of the subject and also an "L" score which indicates the tendency of the subject to falsify his responses to the anxiety items in the scale.

In school achievement, complete test data were available for only 18 of the pairs. For this small group, the test was not analyzed by subscores. When the composite scores were compared, no differences were observed between pairs.

Protocols for the Rosenzweig P-F Study were analyzed along somewhat different lines, since the primary justification for use of this instrument was to enable us to get at larger, more complex units of personality than could easily be subjected to a relatively simple numerical scoring. It was also desirable to have the initial scoring and analysis of the subjects' records done by someone not involved in nor familiar with the matched pair arrangement. For this reason, the Rosenzweig protocols were submitted to an independent consulting psychologist,[3] particularly trained in the use of this instrument, for scoring and analysis. The consultant prepared a scoring summary and a brief written report of each subject's tendencies in response to frustration situations.

The written summaries were next presented to the consulting psychologist in pairs, without identification, and she was asked to rate each pair with respect to which child's record revealed a more "mature," more "integrated," or more "healthy" response to frustration situations. Of the 29 pairs, 12 of the ratings favored group A children, while 17 favored those in group B.

Under the null hypothesis, we would predict that half of the ratings would favor each group. A simple sign test was used to determine the probability of variations as large as those observed. Since we did not predict the direction of differences, a two-tailed test of significance was in order. The probability of variations as great as those observed was found to be .45; thus, there was no basis for rejection of the null hypothesis.

A more detailed analytical approach to the Rosenzweig scoring summaries was next undertaken in an effort to ascertain whether the written reports, necessarily brief and somewhat general, were masking real differences between pairs. Such differences might take the form of specific variations in one or more of the nine scoring factors described in detail by Rosenzweig (6). Briefly, the protocol is scored for each of three directions in which aggressive reactions are made (extrapunitive, intrapunitive, and impunitive) and for each of three types of reaction (obstacle-dominant, ego-defensive, and need-persistent). It is the possible combinations of reaction types with directions of aggression which results in the nine possible scoring factors.

---

[3] We should like to express grateful appreciation to Janet Hirsch for her cooperation and assistance in this study.

The form of our data, with protocols available on a matched-pair basis, suggested the possibility of a kind of analysis which would allow us to examine the interaction effects of groups, reaction types, and directions of aggression simultaneously. The basic question we sought to answer through such an analysis was, of course, whether reaction types and direction of aggression are in any way affected by variation between the two groups of subjects.

A form of the chi square technique was employed for this analysis, results of which are summarized in Table 4. In that table, "Rows" refers to the three directions of aggression, "Columns" refers to the three reaction types, and "Groups" refers to the groups of subjects. It may be inferred from the table that neither the reaction type nor the direction of aggression was in any way affected by the presence of the two groups on a matched-pair basis.

TABLE 4

SUMMARY OF ANALYSIS OF ROSENZWEIG PICTURE-FRUSTRATION STUDY PROTOCOLS

*Variable Combinations**	χ^2	*df*	*p*
Rows × Groups	3.96	2	$\cong .15$
Columns × Groups	1.64	2	$\cong .45$
Rows × Columns	91.35	4	$< .001$
Rows × Columns × Groups	4.43	4	$\cong .35$
Total	101.38	12	

*In the computation of chi square, "Rows" refers to the three directions of aggression (extrapunitive, intropunitive, impunitive); "Columns" refers to the three reaction types (obstacle-dominance, ego-defense, need persistent); and "Groups" refers to the two groups of children.

An interesting sidelight is the highly significant dependence of direction of aggression and reaction type (Rows × Columns) on each other. In general, this dependence seemed to have the form of relatively more impunitive behavior on the part of the obstacle-dominant reaction type, with relatively more extrapunitive behavior on the part of the ego-defensive and need-persistent reaction types. This finding was not pursued, however, since our major concern was with the groups of children both of which showed such trends in about equal degree.

Another type of rating is also obtained on the Rosenzweig data. It is called the "Group Conformity Rating" and indicates the tendency to give more or less popular responses to frustration situations. There were no differences between the two groups with respect to that rating.

Discussion

In this study, scores in the tests of achievement and personality generally favored the control children. In only one case, the Personal Adjustment score of the California Test of Personality, did the difference approach the .05 probability criterion. It would be hazardous to try to explain this only or even largely on the basis of the "experimental" variables. It is conceivable that scores in the California test are in some way associated with one or more, or a combination of, the matching variables in which we achieved limited success. It would be of more than theoretical interest in the professional field of adoption work to determine, for example, the effects of age of adoptive parents on the personality development of children.

For purposes of this presentation, the "unusual mothering conditions" described at the outset could be assumed to include the early discontinuities associated with being placed in a home management house, being cared for there by from 20 to 30 mother figures for a period averaging over five months, then being placed briefly in a foster home, and finally being adopted into a private family. In addition, however, it seems reasonable to assume some further discontinuities in the over-all development of a sense of personal identity, based on the fact of adoption *per se*. One of the obvious limitations of this study is that it does not differentiate between these two types of discontinuities.

For practical purposes at the moment, this may be a somewhat academic question since it has not been demonstrated that real influences on personality are associated with the complex of experiences in the home management house and later adoption. However, the type of control exercised in this study, where no major differences were found to exist, leads us to question whether the same thing would hold true had control subjects been selected during infancy. Presumably, if there are basically distorting or handicapping effects resulting from maternal deprivation, these might well manifest themselves in some of the very things which were used here as criteria for selection of comparison subjects. The most pertinent of these factors, probably, is intelligence. It might be argued that the procedure we have followed, since it meant selecting subjects for comparison whose intelligence was limited to the same level of functioning as that of the respective partner in the adoptive group, might have had an important leveling effect on other aspects of personality, such as those measured by the California Test of Personality, the Manifest Anxiety Scale, and the Rosenzweig P-F Study.

It would seem that the only real answer to that question would have to come from longitudinal studies beginning with the facts of discontinuities in infancy, following subjects over significant periods of development and making comparisons at critical stages with control subjects. One such study (5) is currently in progress.

SUMMARY

Maternal deprivation in infancy should reveal itself in discriminable variations in personality during later childhood, as well as when the unusual mothering occurs. This study was concerned with the long-range effects of an unusual, noncontinuous mothering experience during infancy, a kind of mother-child relation which departs radically from the normal familial patterns of our culture for the first year of life. The problem is whether this constitutes a form of deprivation.

Subjects were 29 children, ages 8 to 17, who had lived in a college home management house during infancy and were subsequently adopted. In the home management house they had been cared for by a large number of mother figures, in serial fashion, experiencing a marked discontinuity in mother-child relations during the first year.

Subjects were paired with children from the same communities, attending the same schools and in the same classes. Pairs were also matched individually on sex, age, and intelligence. Matching for age and education level of parents, family economic level, occupation of father, and number of children in the family was only partially successful.

Subjects completed the California Test of Personality, the Children's Form of the Manifest Anxiety Scale, the Rosenzweig P-F Study, and Iowa Every-Pupil Test of Basic Skills. Statistical comparisons were made, in keeping with the matched-pair design.

The Personal Adjustment score in the California test favored the children who had not been subjected to unusual mothering. Statistically, the difference approached significance. However, this trend was not supported by other findings.

In selecting the tests used, attempts were made to focus on a variety of dimensions of personality processes: school achievement, personal and social adjustment, anxiety level, and response to frustration. In none of these variables could differences be attributed to the factor of discontinuity of mothering in early childhood.

REFERENCES

1. CASTANEDA, A., MCCANDLESS, B., & PALERMO, D. S. The children's form of the manifest anxiety scale. *Child Develpm.*, 1956, 27, 317-326.

2. Gardner, D. B., & Swiger, M. K. Developmental status of two groups of infants released for adoption. *Child Develpm.,* 1958, 29, 521-530.

3. Gough, H. G. A short social status inventory. *J. educ. Psychol.,* 1949, 40, 52-56.

4. North, C. C., & Hatt, P. K. North-Hatt scale to determine relative occupational prestige. *Opinion News,* 1947, 9, 3-13.

5. Pease, D., & Gardner, D. B. Research on the effects of noncontinuous mothering. *Child Develpm.,* 1958, 29, 141-148.

6. Rosenzweig, S., Fleming, E., & Rosenzweig, L. The children's form of the Rosenzweig picture-frustration study. *J. Psychol.,* 1948, 26, 141-191.

7. Siegel, S. *Nonparametric statistics for the behavioral sciences.* McGraw-Hill, 1956.

LATENCY OF SOCIOMETRIC CHOICE AMONG PRESCHOOL CHILDREN

FRANCES DEGEN HOROWITZ*[1]
Southern Oregon College

The investigation of peer relationships among young children provides one approach to the study of socialization. Marshall (4), in a review of sociometric research, concluded that the main body of the investigations failed to provide evidence for the usefulness of the reported sociometric techniques in predicting free play behavior among preschool children. Subsequently, McCandless and Marshall (3) developed a picture sociometric technique in which each *S* is presented with a board containing pictures of his preschool peers and asked to identify each. He is then asked to choose one child whom he best likes to play with, that child being designated his best friend. It was found that *S* choices correlated significantly with teacher best friend ratings and with actual free play behavior, thus offering evidence of the validity of the technique.

Providing an experimental validation of the picture sociometric, Horowitz (2) has demonstrated that a picture of a best friend has more incentive value than a picture of a neutral peer. In that study it was found that young preschool *S*s (mean age 3-6) remained longer at a lever pulling task which revealed a picture of a best friend than did *S*s whose lever pulling responses revealed a picture of a neutral peer.

---

* Southern Oregon College, Ashland.

[1] The author is indebted to Dr. B. R. McCandless for his critical aid in the conduct of this study. The research was carried out at the Iowa Child Welfare Research Station of the State University of Iowa.

Child Develpm., 1961, 32, 235-242.

When a best friend is defined as "a first choice in a sociometric task," the degree of the relation is left unqualified. Such a friendship may represent a strong and relatively constant relation, or it may indicate a weak relation characterized by occasional mutual activities and play periods.

The purpose of the present study was twofold: (a) to develop a technique based upon the picture sociometric which would yield a latency of response measure and (b) to investigate the relation between response latency and the relative strength of peer relations. It was felt that a latency measure would reflect the decision time involved in making a choice.

The hypothesis tested was that *S*s rated as maintaining a strong or medium peer relation would exhibit shorter first choice latencies than *S*s rated as maintaining a weak peer relation. Differential behavior as a function of age was also of interest.

METHOD

Apparatus

The sociometric task apparatus shown in Figure 1 contained 24 three-sided picture frames, each able to hold a 3 by 5 in. picture. A response button was placed below each frame. A Standard Electric Timer connected to the 27 by 36 in. upright board and to *E*'s hand control button recorded the choice latency. Operation of the hand control button started the timer which was stopped when any of the response buttons was pushed.

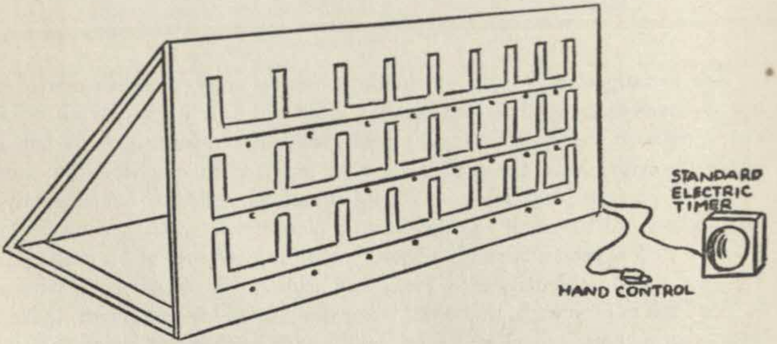

FIGURE 1—Sociometric task apparatus.

A 15 by 19 in. miniature of the sociometric task apparatus containing four picture frames was used as a training task apparatus.

Subjects

*S*s employed were enrolled in the younger (group I) and the older (group II) preschool groups of the Iowa Child Welfare Research Station

Laboratory Preschool. They came from upper middle class families of the professional and managerial class.

Six of the 19 *Ss* in group I were discarded because their choices were made in order of row placement. The mean age of the 12 remaining *Ss* was 3-8, ranging from 3-2 to 4-0. Mean age of the 24 *Ss* in group II was 5-2, ranging from 4-8 to 5-6.

Procedure

Teacher ratings. Two teachers from each preschool group were asked to list for each child in their group the child's best friend. They were also asked to indicate the strength of the friendship; strong, medium, or weak, according to the definitions provided. In subsequent analyses the medium and strong categories were combined, since keeping them separate involved *N*s too small for interpretation. (Hereafter, the strength categories will be referred to as weak and strong.) The teachers were also asked to list three additional friends for each *S* where possible.

Experimental procedure. Each *S* was taken individually from his preschool group to the experimental room by *E* to "play a game." [2] In order to train *S* to push a button rather than point at the pictures, a short training task preceded the sociometric task.

In the training task each *S* was shown the apparatus which contained four pictures of toys (doll, truck, tricycle, and easel painting set). After naming the pictured objects, *S* was told to listen to *E*'s question and then answer it by pushing the button under the picture he chose. *E* said, "Push the button under the picture of the toy you like to play with best." At the conclusion of the question *E* started the timer which stopped when *S* pushed the botton beneath his choice. *E* recorded the choice and latency. *S* was then asked to choose another picture. Then the pictures were changed and the procedure repeated. If *S* indicated his choice by pointing, *E* said, "Push the button to tell me your answer," and then waited until the correct response was made. At the end of the training task all *Ss* were using the buttons to indicate their choices.

The sociometric task immediately followed the training task. *S* was asked to push the button under each picture and to name the child pictured. *E* then said, "Push the button under the picture of the person you like to play with best." Three additional playmate choices were elicited in this manner, substituting the phrase "someone else you like to play with."

For half the *Ss* in each preschool group the pictures were presented in random order A. For the other half, the pictures were arranged in random order B, with the stipulation that a given child's picture did not appear in the corner position in both orders.

---

[2] Author served as *E*. She was not familiar with the teacher ratings until all *Ss* had been run.

TABLE 1

TEACHER RATING RELIABILITY AND TEACHER-SUBJECT CHOICE COINCIDENCE

	RELIABILITY (r)		COINCIDENCE (%)	
	Best Friend	*Strength*	*Weak Ss*	*Strong Ss*
Group I	.28	.78*	58	46
Group II	.62*	.83*	27	61

* Significant at or below .05 level.

RESULTS

Teacher Ratings

The reliability of teacher ratings and the coincidence of teacher-subject choices are shown in Table 1. Strength of friendship was considered as an independent rating, on the assumption that the strength of the best friend relation, as rated by teachers, indicated *S*'s ability to establish associations with peers. The coincidence of teacher-subject choice was here considered without regard to relative order.

Table 2 indicates the percentage of coincidence between teacher and *S* choices, taking order into account. No analyses were performed on these data.

TABLE 2

COINCIDENCE OF TEACHER-SUBJECT CHOICES IN VARIOUS CATEGORIES

	GROUP I		GROUP II	
	Weak	*Strong*	*Weak*	*Strong*
	PERCENTAGE		PERCENTAGE	
S's first choice same as teacher's first choice	0	43	40	56
Teacher's first choice same as one of *S*'s three other choices	60	0	20	11

Sociometric Choice Latencies

The mean latencies of the first sociometric choices are shown in Table 3. No analysis of variance was performed on the data because the requirement of cell proportionality would have necessitated discarding too large a number of *S*s. The value of t for each group was significant at the .08 level of confidence, though the direction of difference was opposite for the two groups.

That is, *S*s in group I rated as maintaining a weak peer relation tended to have longer first choice latencies than *S*s rated as maintaining a strong peer relation. This tendency is reversed in group II.

TABLE 3

MEAN FIRST CHOICE LATENCY
(Time reported in seconds)

	WEAK		STRONG			
	N	*Mean*	*N*	*Mean*	*t*	*p*
Group I	5	5.97	7	3.78	1.84	.08
Group II	15	3.03	9	3.60	1.98	.08

The mean response latency for each *S*'s four choices was computed and a group mean calculated (*see* Table 4). In each case the value of *t* was significant at less than the .02 level of confidence, though the direction of difference in the two groups was opposite. The results, using mean response latency, thus paralleled those from the first choice latency analysis.

TABLE 4

MEAN LATENCY OF FOUR CHOICES
(Time reported in seconds)

	WEAK		STRONG			
	N	*Mean*	*N*	*Mean*	*t*	*p*
Group I	5	5.38	7	3.32	2.30	$<.02$
Group II	15	2.95	9	3.38	2.53	$<.02$

DISCUSSION

Technique

One purpose of the present study was the development of a technique for measuring the latency of a sociometric choice. When presented with a group of pictures and asked to make a choice, pointing is the dominant response among preschool *S*s. Thus, to obtain a measure of first choice latency with a button pushing response, training was necessary. Since all *S*s used the correct response mode on the first sociometric question, the training task seemed adequate.

For all groups first choice latency was longer than the mean latency of four choices, suggesting that, in the series of questions, either *S*s became increasingly familiar with picture placement or learned to anticipate the questions, seeking their choices between and during questions. These factors may have differentially influenced *S*s rated as strong and those rated as weak. If so, there remains to be explained why the influence should have been exerted in one direction for younger *S*s and in the opposite direction for older *S*s.

The technique seems useful, but the apparatus might be modified to include a standard hand rest for a starting position. Incorporating this feature might possibly reduce the variability of the data. However, the size of the board makes central placement of the hand rest difficult.

Teacher Ratings

In three of the four cases, the reliability of teacher ratings was low but significant, with best friend designations showing lower reliability than strength designations. However, strength ratings were not independent, since teachers rated the relation of the best friend and *S*. The assumption involved in considering strength ratings independently is that they were representative of the general intensity of peer relations *S*s were able to establish. This assumption needs further investigation.

Validation of teacher friend designations is found in the percentage of teacher-subject choice coincidence (*see* Table 1). Low percentages would be expected in view of the low reliability for best friend designations. In addition, it would be expected that, given weak friendships, teachers would find it more difficult to list friends than in cases of strong friendships. It is interesting to note that the lowest percentage is found for group II weak rated *S*s.

Choice Latency and Strength of Peer Relation

The main hypothesis was that *S*s rated as maintaining a weak peer relation would exhibit a longer first choice latency than *S*s rated as maintaining a medium or strong peer relation. The results indicated that the direction of the differences between the means for the two groups was opposite. The behavior of the younger *S*s supported the hypothesis, while that of older *S*s was contradictory. Thus, while weak rated *S*s at the younger age level took longer to make their first choices, weak rated *S*s at the older age level made their first choices more quickly.

Two factors might have operated to produce the results. Adults may encourage and reinforce strong peer relations, viewing the socialization of the child as a sign of normal development. As the child becomes more aware of the importance of peer relations, a lack of friends may be accompanied by an amount of anxiety. If the sociometric task elicited anxiety for older *S*s rated as weak in a peer relation, and if anxiety heightens drive, a shorter first choice latency would be expected. More random choices might also result, as there would be less time to think and search out a particular picture. Older *S*s rated as weak did have the lowest percentage of choice coincidence with teacher choice (27 per cent). The results could also be due to the fact that older *S*s rated as strong might have looked more carefully to find their best friends. Choices of strong rated older *S*s showed the highest coincidence with teacher choice (61 per cent). Table 2 indicates that strong rated older *S*s also had the highest percentage of coincidence of first choices with teacher best friend choices, providing some support for this latter explanation.

For the older child, given a greater number of friends, choosing one from among them is a more complex task than when there are a few strong friends since, in a sense, choosing one involves rejecting others. Hence, the approach-avoidance conflict is greater and response may be slowed for two reasons: complexity of task and intensity of conflict.[3] This may lessen the pull of the best friend choice for the *S* who maintains strong friendships, yet interfere little with the choice of the *S* who maintains weak friendships. For the younger child, not aware of the implications of his choice, the hypothesis as tested is working.

Analysis of the mean latency for the four choices yielded results significant at the .02 level of confidence, paralleling the results for the first choice data. A mean latency might provide a more stable measure of each *S*'s latency, thus reducing group variability. Since it corroborated the findings for the first choice latency measure and seemed more stable, it might be the preferred response measure in this type of task.

The present study suggests some further invesigations into peer relations using the modified picture sociometric technique described here. Of particular interest would be further study of the assumption that the weak or strong best friend relation represents *S*'s general ability to establish social relations with peers. Work in this area may provide further clues to the variables affecting the socialization process.

SUMMARY

For each child in two preschool groups teachers listed a best friend, rated the friendship as strong, medium or weak, and indicated three additional friends. Twelve younger and 24 older preschool *S*s were then trained to push a button to indicate a picture choice and were asked to make four friend choices from a modified picture sociometric board. It was hypothesized that *S*s rated as having a weak peer relation would exhibit longer first choice response latencies than *S*s rated as having a medium or strong peer relation. Combining medium and strong rated *S*s, the results supported the hypothesis in the case of younger *S*s, but were antithetical for older *S*s. Factors which might account for the results were discussed and the technique was evaluated.

---

[3] If conflict generates anxiety, thus heightening drive, in this case it would be expected that the increased drive would be reflected in the vigor of the response rather than in an increase of random choices as was suggested for weak rated older *S*s [*see* (1)].

References

1. Finger, F. W. Quantitative studies of "conflict": variations in latency and strength of the rat's response in a discrimination jumping situation. *J. comp. Psychol.,* 1941, 31, 98-127.
2. Horowitz, F. D. Incentive value of social stimuli for preschool children. Unpublished doctoral dissertation, State Univer. of Iowa, 1959.
3. McCandless, B. R., & Marshall, H. R. A picture sociometric technique for preschool children and its relation to teacher judgment of friendship. *Child Develpm.,* 1957, 28, 139-147.
4. Marshall, H. R. An evaluation of sociometric social behavior research with preschool children. *Child Develpm.,* 1957, 28, 131-137.

EFFECT OF HAND-MOUTH CONTACTING ON NEONATAL MOVEMENT[1,2]

JOANNA P. WILLIAMS* *and* WILLIAM KESSEN
Yale University

Contact between the hand and the mouth in the infant has received considerable theoretical attention. The great frequency of occurrence of hand-mouth contact in infants and the continuation of the behavior in the finger-sucking of older children suggests that it is a response of some importance. The fact that this response is intimately related to vital consummatory acts suggesting the possibility that its elicitation or execution may involve some rewarding, relaxing, or "satisfying" effects.

Hoffer (3) has argued that the effectiveness of hand-mouth contact may be due to the fact that it produces a reduction in "tension," a theoretical construct related to "drive," as used, for example, by Hull (4), or "libido," as developed by Freud (2). Some of the behavior patterns found by Wolff (10) in his systematic observations of newborns indicate that such a theoretical formulation may be fruitful.

The general relation between hand-mouth contact and tension, i.e., the correlation between the two variables, that is implied by the theory is not clear. It would seem reasonable, however, that, if hand-mouth contact serves to reduce tension, an effect would be seen in close temporal proximity to the contact. It is expected that tension would be reduced after the infant has made a contact and that tension would rise again as time since the previous contact increases.

Both theoretical and empirical considerations suggest that activity level is a suitable empirical interpretation for the construct of tension. Restlessness is commonly considered to be an index of distress and tension (7, 9), and the theoretical formulations imply a positive relation between degree of tension and activity level. Empirical studies attest to the validity of such an assumption (8, 10).

The present study was conducted to provide evidence relevant to these issues by means of an analysis of the relation between hand-mouth contact

---

* Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut.

[1] The research reported here was supported in part by Grant M-1787, U.S. Public Health Service.

[2] The writers are indebted to David R. Williams for his generous assistance.

Child Develpm., 1961, 32, 243-248.

and activity. It is possible that some mouthing behavior beyond mere stimulation of the mouth region by the hand might be required for a reduction of tension. Wolff (10), for example, suggests that hand-sucking and mouthing before meals may possibly serve to discharge "need tension." For this reason, an analysis of the effects of "long" contacts and of "short" contacts has been made separately.

METHOD

Subjects

The subjects were 10 healthy, full-term newborns, observed in the nursery of Grace-New Haven Hospital. Five were male, and five were female.

Apparatus

A 16-mm. motion picture camera with wide-angle lens was placed in a fixed position approximately 39 inches above the infant. An observer recorded frequency and duration of hand-mouth contacting on an Esterline-Angus operations recorder. The observer's silent telegraph keys also operated a signal light (shielded from the subject) which recorded on the film when a contact was judged.

Procedure

Each subject was observed at 11:00 A.M. on each of his first five days, just before his noon nursing period. The infant was placed on a covered foam rubber pad, dressed in diaper and shirt. A five-minute observation period was followed by five one-minute filmed observations, each one minute apart. In each observation the camera ran until it was completely unwound; thus, the first 30 to 40 seconds of a particular observation were filmed. Inter-observer reliability on recording frequency and duration of hand-mouth contact was .96. Two observers were always present; the second recorded other behavior, not analyzed in the present study. The entire 10 minutes of observation were completed without interruption.

Film Analysis

Observations 1, 3, 5 of days 1, 3, and 5 were analyzed for the present study. Only sessions in which at least one hand-mouth contact occurred were used; this included 85 per cent of the data. The technique has been reported in detail by Kessen, Hendry, and Leutzendorff (6). The present analysis is a modification which affords a measure of activity during any one-second period. For each session 15 frames of the 480 to 640 frames (30 to 40 seconds at 16 frames a second) on the film strip were selected by means of a table of random numbers; these 15 frames served as the beginning-points of the 15 one-second periods for which an activity measure was obtained. The displacements of four standard points of the body from the beginning-point frame to the one 16 frames beyond (i.e., one second) were totaled to determine the activity level for that particular second. For each "activity point" the elapsed time from the last hand-mouth contact as well as the time from the next contact on the film were recorded. When a contact occurred at any point during the 16 frames included in an activity point, no measure was taken. Thus, all activity points obtained are before or after a contact, never during a contact. Inter-observer reliability on this task was .99.

RESULTS

A total of 890 "activity points" were scored; 369 of these fall at some point after a hand-mouth contact, and 461 fall at some point prior to a recorded contact. From inspection of the distribution of contact-duration scores, it was decided to set the arbitrary criterion of a "long" hand-mouth contact at 16 frames or one second duration. It seemed reasonable to assume that a contact of one second or longer was enough for an effect of mouthing to take place. The number of sample points falling into the long and the short categories is approximately equal, so that the sensitivity of the test for either category has not been impaired by this cut-off point.

Significant individual differences in activity were found (within-subject *MS* was 5.63; $df = 9$ and 479; $F = 22.14$; $p < .01$). The score used for each subject was the mean of all his activity scores during a period of eight seconds on either side of a contact; this included about 75 per cent of the data. Mean length of hand-mouth contact for each subject was computed by dividing his total duration of contact by the frequency of contact, totaled over all the observations for that infant. This measure was chosen rather than simple frequency or duration of contact because of the variable amount of film available for each infant. There were no significant differences between subjects. The correlation between mean length of contact and activity was $-.19$, not significant at conventional levels.

Because of the significant individual differences, the complete set of activity scores for each subject was transformed into standard scores in order to equate mean differences. Each score was arbitrarily increased by 4.0 to eliminate negative scores and to facilitate computation.

Analysis of variance showed a significant variation in activity over days: days 3 and 5 were not significantly different from each other; day 1 was significantly different from both days 3 and 5 ($t = 2.40$ and 2.72, respectively).

To assess the effects of hand-mouth contact on activity level just before and just after contact, scatterplots were made of activity as a function of time from the nearest hand-mouth contact. In order to determine the degree of relation between activity and time from hand-mouth contact, Pearson *r*s were computed. Since the present analysis clearly designates activity as the dependent variable and time from contact as the independent variable, the regression of activity on time from contact was also determined. The slope of the regression line indicates how great a change in activity accompanies a given change in time (1).

The statistical significance of an r and its corresponding regression line (b) are the same (1). For both long and short contacts, there was no significant relation between activity and time from hand-mouth contact in the case of time preceding contact. With respect to time following contact, rs of $-.16$ and $-.21$, and bs of $-.0018$ and $-.0020$, all significant at the 5 per cent level, were obtained for long and short contacts, respectively; that is, *S*'s activity decreases very slightly as the time since he last made a hand-mouth contact increases. On the basis of inspection of the scatter diagrams, no test was made of nonlinear trends in the data.

Because there was a significant day variation, however, these statistics were recomputed for days 3 and 5 only. None of the slopes or correlations was significant, when day 1 was omitted from the analysis, nor were any of the results for day 1 alone significant.

There was no significant difference between mean activity level during the first five seconds preceding contact and the first five seconds following contact, either for long contacts or for short ones; that is, activity was not lower just after contact than it was just before a contact. Also, when day 1, and days 3 and 5 were analyzed separately, there were no significant differences.

DISCUSSION

No over-all relation of hand-mouth contact and activity through the lying-in period was shown. This may be due to the fact that this study, not designed primarily to test the gross correlation between these measures, provides an insensitive test; however, it would appear that any effect, if present, is small. The finding that over-all activity level is lower on the infant's first day of life than on his third and fifth is in agreement with other studies concerning developmental changes in activity, as in (5), for example.

The results of the test of the effect more specific to the time of the contact were also not significant. There was no mean difference between activity level during the five seconds before a contact and during the five seconds following a contact, for long or short contacts. Five-second periods before and after contact were chosen because it was felt that any reduction in activity resulting from a hand-mouth contact would be greatest immediately following the contact.

Activity did not rise either as a function of time preceding the occurrence of a hand-mouth contact or as time passed following a contact. Rather, when the data of all three days are pooled, there is a small and significant decrease in activity as time following a contact increases. The fact that such small relations as were found reached significance suggests that this was indeed an adequate test. However, the significance test of the *r* indicates only its reliability and not the magnitude of the relation between the two variables. A very small correlation based on a sample of large *N* often reaches statistical significance, even though for purposes of interpretation the effect is so small as to be trivial. Certainly for the purposes of the present analysis, the slopes of —.0018 and —.0020, even though reliable, are probably not psychologically meaningful.

One limitation on the present finding is that this study ignores the possibility that the only effective hand-mouth contact is, for example, five or six seconds long. This possibility was not considered because of the infrequent occurrence of contacts of such a length. Indeed, the proportion of contacts this long is so small that, if these were the only effective ones, it would seem likely that it was not the hand-mouth contacting alone, but an additional factor present during the longer contacts, that accounted for the variance in activity.

It is concluded that the results of the present study do not support the hypothesis that hand-mouth contact plays an important role in influencing newborn behavior and that, to the extent that activity is related to tension, the hypothesis that hand-mouth contacting functions as a tension-reducing mechanism is not confirmed.

SUMMARY

The relation between hand-mouth contacting and activity, measured by a film analysis technique, was studied in an investigation of the hypothesis that hand-mouth contacting is tension-reducing in the newborn. The present study utilized filmed observations of days 1, 3, and 5 of the lying-in period for 10 newborns.

No significant differences were found between mean activity level during the five seconds preceding a hand-mouth contact and that of the five-second period immediately following a contact. Regression of activity on time preceding a contact was not significantly different from zero. There was a significant but very small decrease (opposite in direction to expectation) in activity as time following a hand-mouth contact increased. When the data for day 1 were eliminated from the analysis, none of the relations was significant; nor were the results of an analysis of the day 1 data alone.

The implication and generality of these findings were discussed.

REFERENCES

1. EDWARDS, A. L. *Statistical methods for the behavioral sciences.* Rinehart, 1956.
2. FREUD, S. Three contributions to the theory of sex. In A. A. Brill (Ed.), *The basic writings of Sigmund Freud.* Random House, 1938. Pp. 551-629.

3. Hoffer, W. Mouth, hand, and ego-integration. *Psychoanal. Study Child,* 1949, 34, 49-55.

4. Hull, C. L. *Principles of behavior.* Appleton-Century, 1943.

5. Irwin, O. C. The amount and nature of activities of newborn infants. *Genet. Psychol. Monogr.,* 1930, 8, 1-92.

6. Kessen, W., Hendry, L. S., & Leutzendorff, A. M. Measurement of movement in the human newborn: a new technique. *Child Develpm.,* 1961, 32, 95-105.

7. Lewin, K. Behavior and development as a function of the total situation. In L. Carmichael (Ed.), *Manual of child psychology.* (2nd Ed.) Wiley, 1954. Pp. 918-970.

8. Marquis, D. P. A study of frustration in newborn infants. *J. exp. Psychol.,* 1943, 32, 123-138.

9. Mittlemann, B. Motility in infants, children and adults. *Psychoanal. Study Child,* 1954, 9, 142-177.

10. Wolff, P. H. Observations on newborn infants. *Psychosom. Med.,* 1959, 21, 110-118.

SOCIOECONOMIC VERSUS FAMILY MEMBERSHIP STATUS AS SOURCES OF FAMILY ATTITUDE CONSENSUS

ROBERT A. DENTLER*
University of Kansas

and J. GLENN HUTCHINSON
University of Illinois

Students of family behavior frequently assume that membership and interaction within one's family group contribute heavily to the shaping of an individual's attitudes, particularly on family-related issues such as the expression of love, companionship, and domestic authority. It seems reasonable to expect that membership in families with distinctive values, standards of conduct, and interaction patterns influences very materially the attitudes of offspring. It is equally plausible to assume that husbands and wives will tend, through extended interaction, toward agreement on family issues.

The aim of this study was to test empirically the extent to which family members do in fact share *common* family-related attitudes. Family membership may influence attitudes in ways that do not result in consensus among members. We assumed, however, that the consensus on salient issues is a satisfactory initial test of family influence and that family group consensus is a problem of interest in itself.

One approach to the problem of the relevance of the family as a source of common attitudes is to contrast the attitudes of family members taken as natural groups with the attitudes of the same individuals regrouped into similarly composed, yet artificially assorted, units. In seeking to learn whether family members had more in common with each other than with other families, the chief aim of our study was to contrast family membership with socioeconomic status.

HYPOTHESES

If families are influential in determining the attitudes of members on family-related issues, then it is reasonable to expect that between two groups (the one a natural kinship group, the other a randomly devised or artificial

---

* Bureau of Child Research, University of Kansas, 9 Bailey Hall, Lawrence.

Child Develpm., 1961, 32, 249-254.

group composed of a father, mother, son and daughter—each drawn from a different kinship group) attitude consensus will be significantly greater within the family group.

Secondly, if the groups are matched by socioeconomic status, it seems equally reasonable to expect that the differences in amount of attitude consensus from status level to level will be greater between the natural families than between the artificial groups. This hypothesis involves testing a secondary assumption, namely that family-related attitudes are bound to vary by socioeconomic level, insofar as the family transmits that portion of the culture accessible to the social stratum and groups in which the parents find themselves. Thus, differences in amount of consensus between groups should be increased when both family and socioeconomic status are controlled.

PROCEDURE

Responses of husbands, wives, and their adolescent children to a battery of 10 questions in the Burgess Middle Years of Marriage Study served as the data for this research.[1] Seventy-five families with one or two adolescent children were selected for analysis.

The 75 families were assigned to three socioeconomic categories, henceforth termed classes, on the basis of the following criteria relating to husbands:

	Number	*Annual Income*	*Occupation*	*Education*
Class I ..	32	$11,500 & over	Professional Executive	16 years & over
Class II ..	21	$ 8,500 maximum	Lower managerial Clerical, sales	10–14 years
Class III ..	22	$ 7,200 maximum	Skilled workers Service workers	12 years maximum

After assigning the families to a class, artificial three- and four-person groups were devised by assorting randomly the subjects within each class. No husband, wife, or adolescent child was sorted into a group containing another member from the same real family.

In the course of home interviews, each respondent had been asked in private to indicate his or her personal opinion of the importance of each of the following items for "the ideal marriage": [2]

---

[1] A follow-up study among the couples participating in the Ernest W. Burgess and Paul Walling research report, *Engagement and Marriage* (Lippincott, 1953). Support for portions of this research, including the present report, was given by the Rockefeller Foundation and the Grant Foundation.

[2] Five degrees of opinion were shown on the checklist, ranging from "Very necessary for a happy marriage" to "Decidedly not desirable."

Love

1. Husband and wife should frequently express their love for their children in words.
2. Husband and wife should frequently express their love for each other in words.

Companionship

3. Parents and children should be just like pals when doing things together.
4. Fathers should spend part of their free time on weekends with the children.
5. Husband and wife should like the same types of amusement (cards, dancing, theatre, etc.).
6. Husband and wife, if congenial, should take their vacations together.

Authority

7. The husband should "wear the pants."
8. The wife should have money of her own, or should earn her own living by paid employment, and not be financially dependent upon her husband.
9. Children should be held to a strict discipline.
10. Young people should be trained never to indulge in "petting" and "spooning."

Each of the three components of the 10-item battery contains items relating to interaction between husbands and wives and between parents and children. The four items in the "Authority" component were selected as an index of traditionalism in this area on the basis of the opinions of a panel of judges and an item analysis which showed these items to be reliable in discriminating between highly "traditionalistic" and extremely "modern" respondents.

Absolute differences in scores were computed on the items between all pairs of individuals. In the 75 families classified according to socioeconomic level, and in the artificial groups, a set of mean differences was computed. Thus, amounts of disagreement refer to total summed differences of pairs, within family and within random groups, on the battery of 10 questions.

RESULTS

Table 1 disconfirms both hypotheses. The first hypothesis, that consensus within natural families will be greater than consensus within artificial groups, is rejected. The difference is statistically not significant for the 75 families compared with the 75 artificial groups; the probability for the gross comparison indicates no more than chance difference ($p = .52$). Indeed, there is *less* consensus within class II families than in class II artificial groups. The second hypothesis, that interclass differences should be greater between classes for natural families, is also not supported.

More interesting is the positive finding that interclass differences, excluding class I versus class II, are significant at beyond the .02 level for

TABLE 1

COMPARISON OF MEANS OF DIFFERENCES BETWEEN INDIVIDUAL SCORES IN FAMILY AND RANDOM GROUPS BY SOCIOECONOMIC LEVEL

Socioeconomic Level	*Family Group*	*Random Group*	*Critical Ratio* (Family vs. Random)
	INTRACLASS MEAN		
Class I	5.55	5.66	.53
Class II	5.63	5.52	.63
Class III	6.04	6.18	.70
	INTERCLASS CRITICAL RATIO		
Classes I vs. II	.4	.6	
Classes I vs. III	2.4	2.5	
Classes II vs. III	2.3	3.3	

both natural families and artificial groups. These findings may be summarized as follows. While differences in amount of consensus are in part accounted for by socioeconomic status, family membership status accounts for none of the variance.

Apart from the hypotheses, Table 1 suggests that attitude consensus on family-related matters is very significantly affected by socioeconomic status. To investigate whether such attitudes might be considered more a function of socioeconomic status than a consequence of family membership, the 10-item attitude score of each family member was paired with that of every other member. Differences between paired scores were then averaged and tabulated as shown in Table 2.

TABLE 2

MEANS OF DIFFERENCES IN SCORES OF INDIVIDUALS IN PAIRS IN FAMILY GROUPS BY SOCIOECONOMIC LEVEL

Socioeconomic Level	*Total* ($N=75$)	*Husband Wife*	*Father Son*	*Father Daughter*	*Mother Son*	*Mother Daughter*	*Child One Child Two*
			LOVE AND COMPANIONSHIP (6 items)				
Class I	4.5	3.3	4.6	4.4	5.1	4.9	5.2
Class II	4.8	3.2	5.1	4.9	5.0	5.2	6.3
Class III	5.3	3.8	5.9	5.6	5.7	5.6	6.1
			AUTHORITY (4 items)				
Class I	4.0	4.2	4.2	5.9	3.9	5.4	4.7
Class II	4.5	5.1	3.9	3.9	4.6	5.1	4.3
Class III	5.0	4.5	5.2	4.9	4.9	5.0	5.8

This approach allowed us to isolate age and sex role sources of attitude consensus, holding socioeconomic status constant. Means of differences within pairs are presented separately for Love and Companionship and for Authority, since means for the two groups of items differed significantly at the .001 level.

A Kruskal-Wallis analysis of variance, using the classified families, was calculated for each of the two scales. On Love and Companionship, the three classes differ significantly at beyond the .10 level. They do not differ significantly on Authority. Therefore, we shall concentrate on the pattern of differences on Love and Companionship.

The "Total" column in Table 2 shows that amount of over-all consensus between paired family members decreases as class level decreases. There is greater consensus within class I families than within class III, as Table 1 also showed.

The influence of socioeconomic status diminishes when family members are paired by age and sex, however.[3] For example, husbands and wives reflect the highest level of consensus among all possible pairs, regardless of class. In contrast, the two offspring per family exhibit extremely low consensus, especially in class II and class III.

Cell variance is not significant on the Authority-related attitudes, yet what pattern occurs parallels the pattern characteristic of differences in consensus on Love and Companionship. For example, consensus decreases with decreasing socioeconomic status. Husbands and wives are more congruent than other pairs, with a few single cell contradictions. Paired adolescent children disagree rather highly, as on Love and Companionship, but the lowest consensus on Authority occurs for paired mothers and daughters. Interestingly, item analysis indicated that mother-daughter disagreement was accounted for by the more extreme conventionality or conservatism among daughters regarding authority.

DISCUSSION

Differences in the family-related attitudes of family members appear to be more readily accounted for in terms of socioeconomic status of the members and of differences in age and sex roles. Although the status classification used in this study involved placing all family members by the income, occupation, and education of one family member, the father, status differences are manifested as completely when members are artificially grouped as they are when family membership is considered.

The findings might be very different if young families with grade school children were used in place of middle-age parents with adolescent offspring. That is, one would expect the influence of groups (and statuses) external to the family to increase with time; and we have shown that the greatest source of disagreement is the children. Parents are relatively highly congruent, socioeconomic status notwithstanding.

---

[3] Differences between paired members on Love and Companionship is significant at below .10 ($\chi^2 = 9.86$, $df = 5$).

Families doubtless transmit, mediate, and modify the attitudes of their members toward the desirability of expressions of love, companionship, and domestic authority. We would suggest that this process of transmission is so extremely "vulnerable" to the influences of socioeconomic status, age status, and sex role prescriptions, among other influences, that future studies of parental attitudes and socialization should explore in greater detail the pressures against family consensus that are thereby generated. Further research should also explore the differential effects of ethnic, religious, and subcultural variations within, as well as among, family groups.

SUMMARY

Responses of parents and children in 75 families to a battery of 10 attitude questions concerning expression of love, companionship, and authority, which were included in the Burgess Middle Years of Marriage Study, were analyzed. Two hypotheses were tested: (a) attitude agreement would be greater in natural families than in artificially composed groups; and (b) class differences, using an index of socioeconomic status, would be greater between families than between artificial groups.

The families were categorized by father's income, occupation, and education into three classes. Artificial groups were devised by random sorting of individuals into three- and four-person groups containing husband, wife, and children, with each drawn from a different family.

Differences between the scores of paired individuals were summed for each family and each artificial group. A test of significant differences using socioeconomic class means and variances of amounts of difference required rejection of both hypotheses. Family groups did not show significantly greater amounts of agreement than artificial groups as a whole or per class. Differences between socioeconomic classes were significant for both family groups and artificial groups, except between class I and class II.

Using mean differences between pairs for a secondary analysis, it was found that on Love and Companionship attitudes (a) consensus was greatest in class I and lowest in class III and the difference was significant at beyond the .10 level and (b) cross-sex and cross-age pairs differed significantly at beyond the .10 level in amounts of agreement. Husbands and wives were uniformly most congruent in their attitude responses.

Family-related attitudes were therefore viewed as affected differentially by determinants that are *not* unique to family membership status, namely socioeconomic status, age, and sex. It was acknowledged that younger families might exhibit less "vulnerability" to external structural effects.

PREDICTION OF OUTSTANDING PERFORMANCE, DELINQUENCY, AND EMOTIONAL DISTURBANCE FROM CHILDHOOD EVALUATIONS[1]

EMMY WERNER* *and* ELIZABETH GALLISTEL
University of Minnesota

This study of special outcomes presents one way of evaluating the predictive effectiveness of pupil responses to inventories and of teacher's ratings in an eight-year study of children's adjustment in a rural county in southwestern Minnesota (1).

PROCEDURE

In 1950 a large number of tests were given to over three thousand school children in this county, all those from grade 4 through grade 12. Inventories to which the pupils responded covered family attitudes, social responsibility, social maturity, home chores performed, range of leisure time activities, pleasant-unpleasant orientation, and psychoneurotic symptoms. In addition, the teachers rated each pupil on 20 personality characteristics (Personality Profile) and on the Havighurst check list of responsible behaviors (Teacher Check List). On the basis of internal consistency, items from these measures were then selected to form a shorter instrument, the Adjustment Inventory. A new responsibility measure was developed from items of the 1950 measure of social responsibility and the Teacher Check List which discriminated between young people nominated as very responsible and others thought of as irresponsible. The sentence completion measure, the chores liked, and the Personality Profile were also shortened by including only items which had discriminated in an item analysis.

These revised measures, together with a new Family Attitude scale used by the University of Minnesota's Rural Sociology Department, were given

---

* National Institute of Neurological Diseases and Blindness, NIH, Bethesda 14, Maryland.

[1] This study has been supported in part by research grant M-690 from the National Institute of Mental Health, U. S. Public Health Service, and in part by funds from the Institute of Child Welfare, University of Minnesota. The principal investigators are J. E. Anderson and D. B. Harris. A major portion of this report was presented at the 1959 meeting of the Society for Research in Child Development.

Child Develpm., 1961, 32, 255-260.

in 1954 to all school children ($N = 3{,}500$) in grades 4 through 12 in the same county.

In 1958 all youngsters from this population who had become delinquent (146) and all who had received public recognition, special positions, or honors (136) were identified. In addition, social workers, school nurses, the probate judge, and the medical clinic provided the names of 18 young people, 13 girls and 5 boys, who had become emotionally disturbed.

RESULTS

The tentative base rates estimated from these referrals and nominations within the county are low. They are, however, based only on young people in their teens and early twenties. For outstanding performance, they are 32 per 1,000 for females, 28 per 1,000 for males. For delinquency based on the number of *individuals, not* on the number of offenses, they are 15 per 1,000 for females, 85 per 1,000 for males, 20 per 1,000 of these for repeated major delinquencies. For those diagnosed as emotionally disturbed, they are 6 per 1,000 for females and 2 per 1,000 for males. There are two reasons for the small size of the emotionally disturbed group: first, it was impossible to determine how many had sought help through a private physician and/or hospital outside the county; second, most of the young people in our study are still below the age at which severe emotional disorders and the major psychoses commonly have their onset.

The mean scores for each of the special groups on the predictor variables are shown in Table 1. They are stanines or standard scores which permit the comparison of a given youngster in a special group with the average for all pupils in the county of his age and sex. The population mean is 5, and the standard deviation, 2.

Both the outstanding boys and the outstanding girls are differentiated by the largest number of predictors. These young people scored significantly higher than their agemates on measures of intelligence and socioeconomic status, on Family Attitudes, Responsibility, Social Maturity, and the Adjustment Inventory. Their teachers rated them above the average on the Personality Profile and the Teacher Check List.

In contrast, boys and girls in the group who later encountered trouble were differentiated by fewer predictor variables. Those who were to become delinquents, both boys and girls, had significantly lower than average scores on measures of socioeconomic status and on the 1954 Responsibility measure. Delinquent *boys* had, in addition, below average ratings on *all* our teacher instruments and on the 1954 Adjustment Inventory. Boys with a record of repeated major delinquencies (rape, burglary, bodily assault with knife, etc.) had considerably lower scores on more predictors than the other delinquent groups. The teachers' evaluations of the *girls* who became delinquent, though low, were not significant for this size sample. These girls, however, did score significantly lower on the 1950 Family measure describ-

TABLE 1

MEAN STANINE SCORES OF OUTSTANDING, DELINQUENT, AND EMOTIONALLY DISTURBED GROUPS ON PREDICTOR VARIABLES

(Means of all members of sample = 5)

	OUTSTANDING		DELINQUENT		EMOTIONALLY DISTURBED
Predictor Measures	*Boys*	*Girls*	*Boys*	*Girls*	*Girls*
IQ (1950)	6.9*	7.2*	5.2	5.3	4.7
SES Measures					
Sewell (1950 SES)	6.7*	6.7*	4.2*	4.3	4.0*
Personal Data (1954 SES)	7.3*	7.1*	4.1*	3.3*	4.5
Father's Education	6.6*	6.2*	5.1	3.9*	4.9
Mother's Education	6.6*	6.6*	4.9	3.9*	5.0
Teacher Ratings					
Personality Profile (1950)	5.9*	5.9*	3.8*	4.4	3.9*
Teacher Check List (1950)	5.8*	6.0*	4.0*	4.2	5.0
Personality Profile (1954)	6.6*	6.2*	3.8*	4.3	3.4*
Pupil Inventories					
Family (1950)	6.6*	6.2*	4.7	4.1*	5.6
Social Responsibility (1950)	6.0*	5.4	4.5	3.8*	4.8
Responsibility (1954)	6.2*	5.8*	4.2*	4.1*	4.1*
Social Maturity (1950)	5.6*	5.8*	4.8	4.8	4.0*
Adjustment Inventory (1954)	6.2*	6.0*	4.0*	4.6	4.2
Not Significant					
Leisure Time Activities (1950)	4.7	5.0	4.9	5.8	4.9
Psychoneurotic Indicators	5.1	5.2	4.9	5.0	5.6
Home Chores (1950)	4.7	4.5	5.1	4.8	4.6
Sentence Completion (1954)	5.2	4.8	4.9	4.4	4.4
Likes (1954)	4.5	4.4	4.8	4.8	5.4
Family Attitudes (1954)	5.0	5.1	4.7	5.3	5.1

NOTE.—The number of cases varies as follows: outstanding boys 40–61, girls 45–69; delinquent boys 80–97, girls 13–18; emotionally disturbed girls 7–16. There were not enough disturbed boys to test for significance of differences.

* Significant at the .05 level or better.

ing home practices having to do with trust, sharing decisions, and closeness.

The emotionally disturbed on whom we have information have significantly lower scores on a measure of rural socioeconomic status. This might, however, have been an artifact of our sampling since we used agency referrals and the county clinic. Table 1 shows only the scores for the girls since the boys' group was not large enough to constitute a sample adequate for statistical analysis. The disturbed girls scored low on the 1954 Responsibility measure and a measure of personal habits, reflecting social maturity. Their teachers rated them significantly lower than their classmates on the Personality Profile, both in 1950 and in 1954.

None of the special groups differed significantly from the population average in the number of chores performed at home and the number of leisure time activities. Fears and worries from the Psychoneurotic Inventory did not predict membership in any of the special groups, not even the emotionally disturbed. We should remember, however, that these tests were taken in middle childhood or early adolescence and that much research needs still to be done to isolate the signs predictive of psychoneurotic behavior in children in contrast with those known of adults.

The 1954 Family Attitude measure did not significantly differentiate any of the special groups, though the 1950 Family measure did. The 1950 Havighurst instrument contained many items of a *specific* nature covering home practices while the 1954 Family measure asked for a *general* expression of positive or negative attitudes toward family members. This evaluation might be more strongly influenced by the temporary emotional state of the youngster, especially an adolescent trying to liberate himself from home.

The differences in the results of these measures point up some of the problems in the evaluation of the relation between certain areas of early adjustment and later outcome. It may be, for example, that there are a number of psychoneurotic signs which can be reliably pinpointed in childhood, but that the form or content of our questions in this area were inadequate. Similarly, it may be that some way of measuring the affect tone of children's responses to a sentence completion measure could be found which would be significantly related to later adjustment, but, though a great deal of research effort was devoted to developing the scoring procedure used for this semiprojective measure, no significant differences were found in these scores for special group members. Both the Sentence Completion and the Likes measures had been found to have diagnostic value, but they had no significant relation to later outcome. There is some indication that they might be unduly influenced by temporary emotions.

Since IQ and SES had differentiated the special groups from the population average, we wondered whether the other measures were significant only because of their relation to those two variables or whether some of the predictors would differentiate members of a special group from agemates of the *same* intelligence and socioeconomic status. Consequently, we tested for significance of difference on all predictors after control for SES and IQ by careful matching and analysis of variance (Table 2).

The Personality Profile proved to be the instrument which most consistently predicted members of special groups after control for IQ and SES, differentiating outstanding members of both sexes, emotionally disturbed girls (no sample for boys), and delinquent boys. It failed to differentiate the delinquent girls whose sample was quite small. The 1954 Responsibility

TABLE 2

PREDICTOR VARIABLES WHICH DIFFERENTIATED SPECIAL GROUPS AFTER CONTROL OF INTELLIGENCE AND SOCIOECONOMIC STATUS BY MATCHING AND ANALYSIS OF COVARIANCE

	OUTSTANDING		DELINQUENT		DISTURBED
Variables	*Boys*	*Girls*	*Boys*	*Girls*	*Girls*
Personality Profile (1950)	0	0	—	0	—
Personality Profile (1954)	+	+	—	0	—
Teacher Check List (1950)	0	0	—	0	0
Responsibility (1954)	0	0	—	—	0
My Jobs MI (Maturity Index)	0	0	—	0	0
Father's Education	+	0	0	—	0
Mother's Education	0	0	—	0	0

+, significant at .05 level or better *above* the mean of control group.
—, significant at .05 level or better *below* the mean of control group.
0, not significantly different from control group.

measure differentiated both boy and girl delinquents. A few other variables (parents' education, social maturity) were occasionally significant.

Finally, we sought to discover whether these instruments singly or in combination constituted an adequate screening instrument around which a preventive program might be built. Frequency distributions by outcome for all individuals in each stanine on the significant predictors were obtained for the follow-up group graduating in 1957 where we had information for every child. These indicated that combinations of IQ, SES, teacher ratings, and significant pupil measures were not sufficiently discriminating to serve as a screening instrument for each individual child. Though there were no outstanding youngsters in the lower three stanines and no delinquents in the upper three stanines, there were enough of *both* groups in the middle stanines to reduce the predictive accuracy beyond the point of usefulness as an independent screening instrument. This was true for the boys more so than the girls. Both positive and negative outcomes were more clearly separated for the girls.

SUMMARY AND CONCLUSIONS

1. Scores on a number of pupil inventories and teacher ratings of groups of school children later recognized as outstanding, delinquent, or emotionally disturbed were significantly different from the norms for their age and sex. Singly or in combination, however, they did not prove adequate for predicting outcomes for a particular individual. The best predictor instruments would seem more valuable as supplements to IQ and SES in-

formation available in school records than as independent screening instruments.

2. A sizeable proportion of the relation between adjustment measures and later outcomes can be accounted for by intelligence and socioeconomic status.

3. After control for IQ and SES, the single predictor which most consistently differentiated special group members from their agemates was a teacher rating scale, requiring ratings of personality characteristics on a five-point scale. As raters, the teachers had the opportunity to compare characteristic behavior of individual children with their agemates and to sample it in different situations over longer periods of time. The personality dimensions covered by the scale were broad enough to characterize both outstanding, disturbed, and delinquent groups and to be meaningful at each developmental stage—late childhood, adolescence, and adulthood.

4. The measure which most effectively differentiated delinquents of both sexes was a pupil instrument measuring social responsibility. It consisted of short descriptive phrases which had been validated empirically against groups known to be high or low on this specific criterion.

5. Measures reporting the range of recreational activities, the number of chores performed at home, the number of experiences liked, the general "affect" as determined from sentence completion tests, and the number of psychoneurotic symptoms all failed to predict membership in any of the special groups.

REFERENCES

1. ANDERSON, J. E., HARRIS, D. B., WERNER, E., & GALLISTEL, E. A survey of children's adjustment over time. A report to the people of Nobles County. Institute of Child Development and Welfare, Univer. of Minnesota, 1959.

VERBALLY EXPRESSED NEEDS AND OVERT MATERNAL BEHAVIORS

VAUGHN J. CRANDALL* *and* ANNE PRESTON
Fels Research Institute for the Study of Human Development

Research concerned with parent-child relationships and the influence of parents on children's social learning and personality development has gradually elucidated a number of avenues through which parents affect their children's development. Recently, child personality development research has been primarily focused on two major avenues of parental influence—the effects of direct parental reinforcements and the more amorphous impact of parents as learning models for children's imitation and identification. Knowledge of the processes of parental influence is, however, incomplete until parents are also studied as individuals *per se*, as well as parents *qua* parents, and relations are explored between the parents' *own* needs, aspirations, and defenses and the behavior they exhibit with their children.

The present study investigates relations between expressed needs of a group of middle-class women and specific overt behaviors they display toward their children. The investigation focuses on the general question of how, and to what degree, such needs find overt expression in maternal behaviors. The study, in contradistinction to most recent research on parental influence relying solely on *reported* parental behaviors obtained through interviews or questionnaires, is based on data of *observed* maternal behaviors of the *S*s as they interacted with their children in their homes.

METHOD

Subjects

The sample was comprised of 38 white, middle-class women, mothers of families participating in the Longitudinal Study of Child Development of the Fels Research Institute for the Study of Human Development and living in industrial centers, in small towns, and on farms within a 30-mile radius of the Institute in southwestern Ohio.

Assessment Procedures

Assessment contacts were made with the *S*s at the Fels Research Institute and in the *S*s' homes. At the Institute the *S*s were administered, among other tests, the Edwards Personal Preference Schedule (PPS). Their re-

---

[*] Fels Research Institute, Yellow Springs, Ohio.

Child Develpm., 1961, 32, 261-270.

sponses to the PPS were scored in accordance with the PPS manual (6), and the scores for the following needs were used as part of the data of the investigation: *n* aggression, *n* dominance, *n* nurturance, and *n* autonomy. These needs were selected for study as ones which might be especially likely to enter into, and influence, maternal behaviors.

Maternal behavior data were obtained through the Fels Home Visit Program. This program, a part of the regular longitudinal data collection procedures of the Fels Longitudinal Study, provides for periodic observations of mother-child interactions as they occur in the homes of the *Ss*. A home visitor, an experienced clinical psychologist specially-trained for the Home Visit Program, makes at least two observation-visits to the homes of the *Ss* each year. Subsequent to each of these visits, the home visitor rates the *S*'s maternal behaviors using the 30 Fels Parent Behavior Rating Scales (PBRs). When observations cover a mother's interactions with more than one of her children, separate PBR ratings are made in respect to the mother's behavior with each child. For the present study, when PBRs were obtained regarding more than one child in a family, the PBR ratings of the mother's behaviors with *one* of her children were randomly selected; the PBR data for the other child or children were excluded. The PBRs so selected represented ratings of maternal reactions towards children ranging from 3 through 8 years of age. The investigation of maternal reactions toward children of such a broad range of ages might be questioned; certain maternal behavior could differ significantly depending on the age of the child. However, previous research (2) has found negligible differences in PBR ratings as a function of age of child within the age span used in the present study.

The time-consuming nature of the Home Visit Program precluded the possibility of determining PBR interrater reliabilities for the sample of this specific study. However, previous research based on the Fels Home Visit Program and the Parent Behavior Rating Scales has demonstrated acceptable inter- and intrarater reliabilities for the scales employed in the current study (2).

The PBR ratings of observed maternal behaviors used in this study were those obtained during the home visit closest in time to the occasion when each *S* was administered the Edwards PPS at the Fels Institute. The median time elapsing between home visit observations and PPS administrations was five weeks.

Hypotheses

The 30 Fels Parent Behavior Rating Scales were originally constructed to assess a large range and diverse variety of parental behaviors (5). Data obtained with these scales have been subsequently analyzed for major "underlying" dimensions of parent behavior by a number of statistical techniques including syndrome analysis, cluster analysis, and first- and second-order factor analyses (2, 4, 7, 8). These various analyses have identified three—and sometimes four, depending on method of analysis—major dimensions of parent behavior. In a previous study, the present authors found four broad parent behavior dimensions. Cluster analysis was employed, and the criterion for the inclusion of a PBR variable in a given cluster was that the variable must correlate .60 or greater with every other variable in the cluster and less than .60 with all other variables not in that cluster. Parental behavior dimensions identified were: (a) affectional behavior, (b) protective behavior, (c) coercive child controls, and (d) coactive child controls. The first three of these dimensions were selected for investigation in the present study and are listed in Table 1. The intercorrelations of the variables representing these three dimensions obtained in the original study are presented in Table 2. The magnitude of correlations ranges from .60 to .84 and indicates that, while the variables within each cluster intercorrelate highly enough to justify the assumption of an underlying dimension, the correlations within each cluster were not sufficiently high to permit the PBR scales of these variables to be used interchangeably. In none of the correlations was as much as three-quarters of the variance held in common. In the present study, therefore, while data analysis was organized around the three major maternal behavior dimensions investigated, and general hypotheses were stated in respect to these, tests of these general hypotheses were made by ascertaining relations between the *Ss'* PPS need scores and *each* of the PBR variables representing the maternal behavior dimensions under investigation.

TABLE 1

PBR SCALES REPRESENTING THE THREE MAJOR DIMENSIONS OF MATERNAL BEHAVIOR INVESTIGATED IN THE PRESENT STUDY

PBR Scales Representing Affectional Maternal Behavior

Affectionateness toward child: hostile – affectionate
Rapport with child: distant – close
Direction of criticism: disapproval – approval
Intensity of contact: vigorous – inert

PBR Scales Representing Protective Maternal Behavior

Protectiveness: exposing – sheltering
Solicitousness for child's welfare: nonchalant – anxious
Babying: withholds help – overhelps
Child-centeredness of home: child-centered – child-subordinated

PBR Scales Representing Coercive Maternal Control

Severity of penalties: mild – severe
Readiness of enforcement: lax – vigilant
Coerciveness of suggestions: compliance is optional – mandatory

Table 2

INTERCORRELATIONS OF PBR SCALES REPRESENTING THE THREE MAJOR MATERNAL BEHAVIOR DIMENSIONS INVESTIGATED IN THE PRESENT STUDY
($N = 44$ in original study)

PBR Variable	Intercorrelations		
Affectional maternal behavior	*2*[*]	*3*	*4*
1. Affectionateness	.79[†]	.80	.79
2. Rapport		.76	.60
3. Direction of criticism (approval)			.74
4. Intensity of contact			
Protective maternal behavior	*2*	*3*	*4*
1. Protectiveness	.74	.83	.62
2. Solicitousness		.67	.76
3. Babying			.61
4. Child-centeredness			
Coercive maternal control	*2*	*3*	
1. Severity of penalties	.82	.73	
2. Readiness of enforcement		.84	
3. Coerciveness of suggestions			

[*] As an aid to inspections in each maternal behavior cluster, the numbers above the correlations correspond to the numbers given the PBR variable in the "PBR Variables" column. These numbers are *not* the original numbers of the specific PBR scales.

[†] All correlations in Table 1 are product-moment correlations run on normalized data. All correlations presented are significant beyond the .01 level of confidence.

The study explored relations between middle-class women's verbally-expressed needs and their overt maternal behaviors. Operationally, the *Ss*' needs were determined by their Edwards PPS responses; their maternal behaviors were quantified in PBR ratings based on home visit observations of mother-child interactions.

Four general hypotheses were tested. These related the *Ss*' *n* nurturance, *n* aggression, *n* autonomy, and *n* dominance to maternal behavior dimensions. *Hypothesis I:* "Middle-class women's *n* aggression will be negatively related to the amount of overt affection they display toward their children." Parenthetically, it should be noted that hypothesis I, as well as the three hypotheses which follow, is tenable only to the degree that women's verbally-expressed needs find direct expression in maternal behaviors. When internal or external inhibiting forces are present, such needs may not eventuate in relevant actions. Women with aggressive propensities, for example, may not behave aggressively toward their children when internal checks are present such as the anticipation of guilt following aggression and/or when external inhibiting forces exist, e.g., possible criticism from others for hostile maternal behavior. In addition, mothers may report aggressive urges but use persons other than their children as primary targets for their aggression, e.g., husband, acquaintances, or various minority groups. Nevertheless, while some or all of these possibilities may occur, the present study tests the general proposition that middle-class women's verbally-expressed needs will, beyond chance, find behavioral expression in their parent-child interactions. In reference to hypothesis I, it is expected that women who freely acknowledge aggressive urges will express such aggression overtly toward their children by nonaffectionate and rejectant behaviors. Concerning the specific PBR variables representing the dimension of maternal affection (cf. Table 1), it is predicted that, in comparison with *Ss* with low *n* aggression, *Ss* who readily report aggressive desires will: (a) express less overt affection toward their children, (b) maintain less close rapport with them, (c) tend generally to criticize rather than approve of their children's behaviors, and (d) have less intense contacts with their children.

Hypotheses II and III are concerned with women's needs as predictors of their overt protective maternal behaviors. *Hypothesis II:* "Middle-class women's verbally-expressed *n* nurturance will be positively related to the amount of protectiveness they display toward their children." In respect to specific protective maternal behaviors (cf. Table 1), it is predicted that *Ss* with high *n* nurturance will be especially prone to: (a) act protectively toward their children, tending to shelter them from, rather than exposing them to, potential physical and psychological vicissitudes, (b) be solicitous, i.e., concerned rather than nonchalant, regarding their children's welfare, (c) infantalize their children by overhelping rather than withholding help, and (d) maintain a home in which activities are child-centered rather than child-subordinated.

Hypothesis III relates women's *n* autonomy and their protective maternal behaviors. *Hypothesis III:* "Middle-class women who express strong needs for autonomy will exhibit little protective behavior toward their children." The rationale underlying this hypothesis is less straightforward than those of hypotheses I and II. For the two previous hypotheses, it was assumed that the needs investigated would be directly translated into maternal behaviors; *n* aggression would result in rejectant maternal behaviors and *n* nurturance, in protective maternal behaviors. The reasoning behind hypothesis III is more circuitous. In fact, either of two processes may produce the predicted relation between *n* autonomy and maternal protectiveness. First, mothers with strong needs for autonomy and independence might be expected to value similar needs in their children—at least to the degree that these mothers identify with their children and react to them as extensions of themselves. Such mothers might, then, instigate and reward autonomous behaviors of their children and discourage dependent overtures such as help-seeking or emotional support-seeking calling for maternal nurturance.

A second possibility also exists. Mothers with strong *n* autonomy may *not* necessarily value autonomous behavior of their children and, therefore, react nonnurturantly when their children make dependent demands on them. Nevertheless, such women may still exhibit little nurturance with their children insofar as they feel their children's dependent overtures are encroachments on their *own* need for autonomy. If the latter were the case, these women might be expected to be especially likely to have child-subordinating, rather than child-centered, orientations (cf. Table 1). In general, however, in respect to hypothesis III, it is predicted that women with a strong *n* autonomy will, in comparison with women with low *n* autonomy, be: (a) less protective of their children, (b) less babying, (c) less solicitous, and (d) less child-centered.

A final hypothesis relates *n* dominance and coercive maternal controls. *Hypothesis IV:* "Women who readily express a need to dominate others will be especially prone to employ coercive controls with their children." Specifically, it is predicted that *S*s with high *n* dominance, in comparison with *S*s with low *n* dominance, will: (a) evoke more severe penalties when their children do not follow their rules and regulations, (b) more readily enforce the rules they establish for their children's behavior, and (c) use more coercive suggestions with their children regarding what they—the children—should or should not do.

Results

Table 3 presents relations between the *S*s' verbally-expressed needs and their overt maternal behaviors. The data in this table have been organized in reference to the four major hypotheses of the study as well as in respect to the specific PBR variables representing the three general maternal behavior dimensions investigated.

Hypothesis I, that the *S*s' *n* aggression would be negatively associated with the amount of affection they displayed with their children, found general confirmation. Women with high PPS *n* aggression expressed less overt affection toward their children, had less close rapport with them, tended to criticize rather than approve of their children's behavior, and had less intense contacts with their children. In general, then, as represented by the sample of the present study, the verbally-expressed need for aggression of middle-class women *does* find direct behavioral expression in the nature of the affectional relationships these women establish with their children.

Hypothesis II, that the *S*s' *n* nurturance would be positively related to the amount of protectiveness they expressed toward their children, also found substantiation. The *S*s whose Personal Preference Schedule responses indicated a strong general need to nurture others were especially protective of their children. In comparison with the *S*s with low *n* nurturance, in everyday life these women exhibited more protectiveness toward their children,

TABLE 3

RELATIONS BETWEEN SUBJECTS' VERBALLY-EXPRESSED NEEDS AND THEIR OBSERVED MATERNAL BEHAVIORS
($N = 38$)

PPS Variable	*PBR Variable (and direction)*	*r*	*p*
Hypothesis I—Aggression			
	Ready expression of affection	—.40	.01
	Close rapport with child	—.53	.01
	Direction of criticism (approval)	—.54	.01
	Intense contact with child	—.36	.05
Hypothesis II—Nurturance			
	Much protectiveness	.48	.01
	High solicitousness	.50	.01
	Tendency to baby child	.44	.01
	Child-centered orientation	.45	.01
Hypothesis III—Autonomy			
	Much protectiveness	—.25	*ns*
	High solicitousness	—.22	*ns*
	Tendency to baby child	—.23	*ns*
	Child-centered orientation	—.39	.05
Hypothesis IV—Dominance			
	Employs severe penalties	.23	*ns*
	Readily enforces rules	.29	*ns*
	Coercive suggestions	.07	*ns*

were more solicitous of their children's welfare, tended to overhelp rather than underhelp them, and established homes where child-centered orientations predominated.

Hypotheses III and IV received less clear-cut support. In all four correlations concerned with hypothesis III, directions of relations were those predicted; *n* autonomy tended to be negatively associated with protective maternal behaviors. However, only one correlation—that between *n* autonomy and child-centeredness—was statistically significant. All relations concerned with hypothesis IV were in the predicted direction; none was significant.

DISCUSSION

The results indicate that some verbally-expressed needs of middle-class women are directly translated into maternal behaviors. Verbal reports of aggressive urges significantly predict a number of different rejectant maternal behaviors. Similarly, the expression of a general desire to nurture others found behavioral expression in the *S*s' everyday interactions with their children. However, other needs were not, as hypothesized, significant predictors of the *S*s' behavior with their children. Only one of four correlations between the *S*s' *n* autonomy and their protective maternal behaviors was statistically significant. The fact that the significant relation occurred between high *n* autonomy and low child-centeredness suggests—though does not conclusively prove—that the need for autonomy in middle-class mothers is less a determinant of low maternal protectiveness as a way of developing valued autonomous behavior in their children than it is a more direct reflection of egocentric desires by the strong *n* autonomy mothers to be "left alone" by their children. It appears that high *n* autonomy mothers may express less nurturant maternal behaviors simply because they feel their children's dependent demands are encroachments on their *own* autonomy.

The final hypothesis that the *S*s' needs to dominate others would find behavioral expression in the coerciveness of the control methods they employed with their children was not substantiated. Several alternative *post hoc* explanations can be given for the lack of significant relations between variables testing hypothesis IV. None of these explanations, however, can be tested with the data of the present study. One possibility is that the *n* dominance scale of the Edwards PPS is not a predictively useful measure of adult women's need to dominate others (including their children). The PPS, while it is one of the most meticulously constructed of current personality questionnaires (e.g., controlling for social desirability of responses), still—with a few notable exceptions, such as the study by Bernadin and Jessor (3)—awaits construct validation research. In a limited sense, the current study presents evidence for the construct validity of the PPS *n* aggression and *n* nurturance scales. The scales *do* predict the amount of aggression and nurturance mothers express toward their children. It is also interesting to note that middle-class mothers' PPS *n* aggression and *n* nurturance scores are as good, or better, predictors of their overt affectional and nurturant maternal behaviors as are their own self-ratings of their maternal behaviors in these areas. In a previous study (4), the authors constructed self-rating scales of maternal behavior to parallel the same Fels PBR scales employed in the present study. In the previous study, a group of middle-class mothers similar to, and including some of, the *S*s of the present study rated their own maternal behaviors with these specially-constructed mother self-rating scales. These self-ratings were correlated with parallel PBR ratings made by the Fels Home Visitor on the basis of her observations of these mothers' overt interactions with their children. In respect to the general maternal behavior dimension of affection, correlations between the mothers' self-ratings and the Home Visitor's PBR ratings were .59, .33, and .45 for affectionateness, rapport, and direction of criticism, respectively. These correlations are of the same general magnitude as those obtained in the present study between PPS *n* aggression scores of the *Ss* and these same PBR variables (cf. Table 3). In other words, middle-class women's PPS *n* aggression (an assessment of need not specific to mother-child relationships) predicts their overt affectional maternal behaviors as well as do the mother's own self-ratings of these specific maternal behaviors. Moreover, in respect to the prediction of protective maternal behaviors, women's PPS *n* nurturance scores (again not specifically pertaining to mother-child relationships) predict overt protective maternal behaviors *better* than do the mothers' own self-ratings of these behaviors. In the present study, PPS *n* nurturance was significantly correlated with PBR ratings of protectiveness, babying, solicitousness, and child-centeredness; in the previous study, mothers' self-ratings of these same behaviors did *not* predict the degree to which these maternal behaviors actually occurred in everyday mother-child interactions. For some areas of maternal behavior, then, the verbally-expressed needs of women seem to be as good, or better, predictors of their actual maternal behavior practices than are these women's subjective assessments of their own childrearing practices.

Returning to the results of the present study, the fact that the Edwards PPS *n* autonomy and *n* dominance scales were not, as hypothesized, generally predictive of the *Ss*' overt maternal behavior may have been due to a number of factors including: (a) poor construct validity of the scales for the population sampled, (b) poorly formulated hypotheses, and/or (c) a lack of "true relations" between the variables measured. Answers to these alternative possibilities await future research. At the moment, however, the findings of the present study warrant these general conclusions: (a) Middle-class women's verbally-expressed needs for aggression and nurturance clearly predict the degree that these women *overtly* express nonaffectionate and nurturant behaviors toward their children. (b) There is a lesser trend that the *Ss*' *n* autonomy is associated with nonnurturant aspects of their maternal behavior. (c) No evidence was found that the *Ss*' stated desires to dominate others were directly translated into coercive maternal behavior practices.

SUMMARY

This study explored relations between certain verbally-expressed needs of middle-class women (as measured by their Edwards PPS responses) and the overt behavior these women displayed toward their children. Data regarding the latter were obtained from home visit observations of mother-child interactions and subsequent ratings of these maternal behaviors using the Fels Parent Behavior Rating Scales. Four general hypotheses were tested. These were: *Hypothesis I:* "Middle-class women's verbally-expressed *n* aggression will be negatively correlated with the amount of affection they display toward their children." *Hypothesis II:* "The Ss' expressed need to nurture others will be positively associated with the protectiveness they exhibit with their children." *Hypothesis III:* "Women with high *n* autonomy

will be less protective of their children than will women with low *n* autonomy." *Hypothesis IV:* "Women who express strong needs to dominate others will be particularly prone to employ coercive maternal controls in their everyday child-socialization practices." Hypotheses I and II were clearly substantiated. Hypothesis III found limited substantiation; hypothesis IV, none. Implications of the findings were discussed.

REFERENCES

1. BALDWIN, A., KALHORN, J., & BREESE, F. Patterns of parent behavior. *Psychol. Monogr.*, 1945, 58, No. 3 (Whole No. 268).
2. BALDWIN, A., KALHORN, J., & BREESE, F. The appraisal of parent behavior. *Psychol. Monogr.*, 1949, 63, No. 4 (Whole No. 299).
3. BERNADIN, A., & JESSOR, R. A construct validation of the Edwards Personal Preference Schedule in respect to dependency. *J. consult. Psychol.*, 1957, 21, 63-67.
4. CRANDALL, V. J., & PRESTON, A. Patterns and levels of maternal behavior. *Child Develpm.*, 1955, 26, 267-277.
5. CHAMPNEY, H. The measurement of parent behavior. *Child Develpm.*, 1941, 12, 131-166.
6. EDWARDS, A. L. *Manual; Edwards Personal Preference Schedule.* Psychol. Corp., 1954.
7. LORR, M., & JENKINS, R. L. Three factors in parent behavior. *J. consult. Psychol.*, 1953, 17, 306-308.
8. ROFF, M. A factorial study of the Fels Parent Behavior Scales. *Child Develpm.*, 1949, 20, 29-45.

AUTHORITARIANISM, OCCUPATION, AND SEX ROLE DIFFERENTIATION OF CHILDREN

RONALD C. JOHNSON,* CAROL JOHNSON, *and* LEA MARTIN
San Jose State College

As pointed out by Linton (2), sex role differentiation begins for children in our culture long before there are any significant sex differences in potentials for work or for social participation. By the time children are 3 or 4 years old, quite different parental expectations and disciplinary techniques have developed for girls as opposed to boys (5). Since sex role differentiation serves little, if any, real function in early childhood, the amount of role differentiation imposed by the parent would seem likely to show considerable variation.

Few studies have been done in this area. Scheinfeld (6) and Rabban (4) have shown parental social class to be significantly related to the amount of rigidity shown by parents in the delineation of children's sex roles. Lower occupational groups appear to differentiate sex roles earlier and to a greater extent, leaving a smaller area of behavior permissible to both sexes, than do middle socioeconomic groups. A number of explanations for the obtained differences might be advanced. Jobs that are lower on the job hierarchy may be more masculine than those at the upper end (e.g., bulldozer operator vs. auditor) and hence allow boys from lower economic groups a more masculine model with which to identify. Perhaps there are class differences in ideas about what constitutes masculinity and femininity.

Social classes differ, not only in definition of sex appropriate behaviors, but in other ways as well. One area of difference is in authoritarianism, as measured by the California F Scale (1). Among the hypothesized attributes of those individuals scoring high (on the authoritarian end) on the F Scale is an intolerance of ambiguity (1). Lower economic groups score higher on the F scale than middle or upper economic groups (1). Is, then, their more rigidly defined sex role differentiation due, basically, to their higher degree of authoritarianism?

This study is an attempt to determine the relation of socioeconomic characteristics and of authoritarian attitudes to sex role differentiation.

METHOD

Subjects

Fifty-eight college students taking an upper division course in child psychology served as *S*s. Eighteen were parents; 40 were not. Thirty-nine *S*s were female; 19 male.[1] It is not claimed that these *S*s were representative

---

* Psychology Department, San Jose State College, San Jose, California.

[1] Neither marital status nor sex of subject was related to other variables measured in this study.

Child Develpm., 1961, 32, 271-276.

of the occupational groups into which they were categorized on the basis of parental occupation. By the mere fact that they were college students, these *S*s constituted a biased sample. However, they were a relatively homogeneous sample in terms of intellectual ability, a definite advantage since many obtained social class differences may be due, basically, to variation between social classes in tested intelligence.

Measuring Devices and Procedure

Information concerning parental occupation was obtained from the *S*s. Subjects were divided into occupational groups according to the Minnesota Occupational Scale (7). Subjects were also categorized into "entrepreneurial" and "bureaucratic" groups according to parental occupation, following a distinction made by Miller and Swanson (3). Entrepreneurial parents were defined as those who were employed in small sized businesses characterized by a simple division of labor and small capitalization in which upward mobility and income are obtained through risk taking and competition. Entrepreneurs would be such people as independent stock brokers, realtors, shopkeepers, door to door salesmen, auto salesmen, small contractors, contract fruit and vegetable pickers, or even doctors or clinical psychologists unaffiliated with any institution—men dependent chiefly on their own efforts for income and status. Bureaucratic parents, on the other hand, would be members of large organizations, specialized in division of labor and usually highly capitalized. Prestige and income come not from risk taking or competition but from cooperation, development of specialized skills, and seniority. A plant guard, a clerk in a department store, a General Motors executive, or a college professor would fall into the bureaucratic group.

Subjects were administered the F scale and were then presented with a list of 44 children's behaviors compiled by the senior writer, in order to determine the amount of sex role differentiation made by each *S*. All of the listed behaviors were presumably of a type that both girls and boys could undertake them, although a number of them were more characteristic of one sex than the other. *S*s were told:

> Here is a list of things that children aged 6 through 11 do. After each behavior mark down whether you think it to be appropriate for *both* boys and girls, for *boys,* not girls, for *girls,* not boys, or for *neither* sex. For example, if you believe it proper for girls but not boys to cry when hurt, you would write *girls* after the first listed behavior.

RESULTS

Of the 44 behaviors, the number of behaviors rated as appropriate for both sexes by these *S*s ranged from 15 to 44. The behaviors rated, along with the difference in the frequency that they were rated as appropriate for one as opposed to the other of the sexes, are presented in Table 1.

TABLE I

RATINGS OF SEX APPROPRIATENESS OF BEHAVIOR

Behavior	*Difference in Ratings**	*Behavior*	*Difference in Ratings*
Pretend to be father	+52*	Run errands	+ 5
Have paper route	+51	Refuse to obey	+ 3
Play with toy soldiers	+32	Criticize parents	+ 1
Shoot slingshots	+31	Read comic books	+ 1
Play marbles	+30	Ride bikes	0
Build tree houses	+29	Watch T.V. cartoons	0
Play with toy trucks and cars	+28	Play hide and seek	− 1
Fight back when picked on	+25	Cry when hurt	− 3
Play cops and robbers	+23	Paint pictures	− 4
Go to movies alone	+21	Play store	− 4
Help with lawn	+18	Write poems	− 5
Take dares	+18	Be sensitive to feeling of others	− 6
Go around town alone	+14	Learn to dance	−12
Collect insects, bugs	+13	Make own bed	−12
Fly kites	+13	Read love stories	−16
Collect rocks	+10	Jump rope	−29
Go on hikes	+10	Play jacks	−36
Skip school	+10	Help mother cook	−37
Read adventures	+ 8	Play with dolls	−48
Watch T.V. cowboys	+ 7	Play nurse	−52
Collect stamps, coins	+ 7	Pretend to be mother	−55
Start secret club	+ 6		

* Frequency of boy-appropriate ratings minus frequency of girl-appropriate ratings: plus sign indicates more appropriate for boys, minus sign, more appropriate for girls.

Many of these behaviors range far more widely from neutrality than one would expect. Even in this group of college *S*s, a very wide differentiation of sex roles appears to exist. Two kinds of role restriction can be measured: (a) the number of behaviors said to be inappropriate to both sexes—a "general" form of restrictiveness—and (b) the number of behaviors said to be appropriate to only one sex—a restrictiveness built around the degree of differentiation of behavior expectancies for boys as opposed to girls. The variation between *S*s in both forms of restrictiveness was analyzed to determine the relation between restrictiveness, F scale score, and measures of occupational level and type.

Forty-six of the *S*s were reliably (unanimous agreement of three judges) placed into occupational levels I, II, III, V, VI, and VII of the Minnesota Occupational Scale on the basis of parental occupation. They were grouped into levels I and II (professional and semiprofessional), III and V (clerical, small business, skilled and semiskilled labor), and VI and VII (slightly

TABLE 2

OCCUPATIONAL GROUP DIFFERENCES IN SEX ROLE ASSIGNMENT AND F SCORE

	MEAN NUMBER OF BEHAVIORS ASSIGNED				
Occupational Group	*Both*	*Boys*	*Girls*	*Neither*	*Mean F Score*
I & II ($N = 18$)	26.3	7.6	5.6	4.5	3.4
III & V ($N = 18$)	24.3	8.3	6.5	4.9	3.1
VI & VII ($N = 10$) ...	26.0	7.0	5.4	5.6	3.7

skilled and unskilled labor). The role categorizations and the F scores of these three groups are presented in Table 2.

No significant differences exist between these groups on any of the variables measured. One would expect, from the literature, to find differences both in sex role differentiation and in F scale score.

Miller and Swanson (3) found very few significant differences in child-rearing attitudes and behaviors among a random sample of mothers divided in social classes according to the Warner Index (upper-upper, lower-upper, etc.). They did, however, find very considerable differences between mothers when they were divided according to whether they belonged to entrepreneurial or to bureaucratic groups. The data collected in this study were reanalyzed according to this differentiating scheme. Thirty-five *Ss* were reliably (unanimous agreement of three judges) placed, on the basis of parental occupation, in either entrepreneurial or bureaucratic groups. Fifteen *Ss* fell into the entrepreneurial category; 20 into the bureaucratic group. The mean number of behaviors rated as appropriate for both sexes was 23.3 for the entrepreneurial group; 32.6 for the bureaucratic group. A Mann-Whitney test indicated that the differences in ranks between these two samples in the number of behaviors rated as appropriate for both sexes is significant ($p < .05$); the bureaucratic group has a significantly larger number of behaviors that are appropriate and are sexually neutral. No other differences in the ratings of these behaviors in terms of sex appropriatenesss are significant, although, since the entrepreneurial group rated less behaviors as appropriate to both sexes, they necessarily placed more behaviors in the other three categories.

Significant differences existed between the entrepreneurial and the bureaucratic groups in F scale scores as well as in ratings of sex appropriateness of behavior. F scale scores had been obtained for all 58 *Ss*. All *Ss* were divided into three groups—high, medium, and low—according to the third of the distribution into which their F scale score fell. The 35 *Ss* who were reliably categorized on the entrepreneurial-bureaucratic dimension fall into the various thirds of the distribution as is shown in Table 3. Table 3 also

TABLE 3

NUMBER OF APPROPRIATE, SEXUALLY NEUTRAL BEHAVIORS BY F SCORE GROUP AND BY OCCUPATIONAL TYPE

	F SCALE GROUPS		
	Low	*Medium*	*High*
Entrepreneurial			
Number of Subjects	1	4	10
Mean Number of Behaviors	31.0	26.5	21.2
Bureaucratic			
Number of Subjects	12	6	2
Mean Number of Behaviors	35.7	29.8	22.5

gives the mean number of appropriate, sexually neutral behaviors checked by *S*s by F score groups and by entrepreneurial-bureaucratic differentiation.

F scale scores are not evenly distributed between the two groups ($\chi^2 = 16.0$, $df = 2$, $p < .01$), with the entrepreneurial group being significantly more authoritarian. A Kruskal and Wallis H test shows that the "appropriate to both boys and girls" rating is not distributed randomly ($H = 8.25$, $df = 2$, $p < .05$) between *F* scale groups. It is clear that the vast majority of the differences between entrepreneurial and bureaucratic groups in the ratings of these behaviors is related to differing F scores rather than to entrepreneurial vs. bureaucratic occupation *per se*.

DISCUSSION

Within this sample of *S*s (admittedly a relatively homogeneous sample in terms of IQ and probably in terms of values and aspirations) no significant relations are to be found between position on the Minnesota Occupational Scale and either F score or number of behaviors rated as permissible for both sexes. This may be due to the homogeneity of this biased sample in intelligence and/or in values. It may be, too, that social class differences, with class measured by the Minnesota Occupational Scale or the Warner Index, existed at one time, but do not exist to any large degree at the present time, as suggested by the recent finding of Miller and Swanson of no differences in attitudes or behaviors between *S*s categorized according to the Warner Index.

Once *S*s are differentiated into entrepreneurial vs. bureaucratic groups, on the basis of parental occupation, significant differences in both F scale score and in permissible, sexually neutral behaviors are obtained. Table 3 indicates that, if matched for F score, the entrepreneurial-bureaucratic categorization would not, in itself, allow for significant differences in the number of behaviors rated as both permissible and sexually neutral. A major aspect of authoritarianism is said to be intolerance of ambiguity. These data tend to support this proposition. A clear cut differentiation of sex roles seems to be rather closely related to authoritarianism. The entrepreneurial group is significantly more authoritarian (as authoritarianism is measured by the F scale), as one would expect from Miller and Swanson's analysis. It seems likely that this high amount of authoritarianism is manifested by *S*s in clearly defining the sex appropriateness of children's behaviors.

SUMMARY

A college population of *S*s was given the California F (authoritarianism) scale. *S*s rated a number of children's behaviors in terms of their sex appropriateness. Measures of parental occupational level and type were used to form subgroups of *S*s. No significant differences in F scores or in ratings of children's behaviors were obtained between *S*s of differing parental occupational levels on the Minnesota Occupational Scale. When *S*s were divided into entrepreneurial and bureaucratic groups, on the basis of parental occupations, significant differences did appear, with entrepreneurial *S*s scoring higher on the F scale and rating significantly fewer behaviors as being appropriate to both boys and girls. When behavior ratings were divided within entrepreneurial and bureaucratic groups according to F scale score, F score was significantly related to the number of "appropriate to both boys and girls" behaviors. With F scores held constant, only slight differences appear between the ratings of entrepreneurial and bureaucratic groups. Certain implications of these findings were discussed.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* Harper, 1950.
2. LINTON, R. Age and sex categories. *Amer. sociol. Rev.,* 1942, 7, 589-603.
3. MILLER, D. R., & SWANSON, G. E. *The changing American parent.* Wiley, 1958.
4. RABBAN, M. Sex role identification in young children in two diverse social groups. *Genet. Psychol. Monogr.,* 1950, 42, 81-158.
5. RADKE, M. J. The relation of parental authority to children's behavior and attitudes. *Univer. Minnesota Child Welf. Monogr.,* 1946, No. 22.
6. SCHEINFELD, J. *Women and men.* Harcourt, Brace, 1944.
7. *The Minnesota Scale for Paternal Occupations.* (Rev.) Institute of Child Welfare, Univer. of Minnesota, 1952.

JUDGMENT OF SIZE IN RELATION TO GEOMETRIC SHAPE

Betsy Worth Estes*[1]
University of Kentucky

This is an investigation of perceived size comparisons for figures of the same shape and figures of different shape.

Previous investigators have studied the development of simple perceptual tasks, such as the ability of very young children to select the middle size of three shapes (3, 8, 9), and concepts of roundness (4) and of largest size (5). In these studies the authors were concerned essentially with the development of meanings, of concepts.

Many studies have been concerned with judgment of size as related to constancy and thus to learning and concept formation. Some have questioned whether the tasks which have been used are of a simple perceptual nature or whether more complex judgment is involved. Relevant variables which have been investigated are the experimental situation, attitudes of subjects, age, and intelligence. Osgood agrees with Boring (1) that Martius' almost perfect constancy "looks more nearly like judgment than immediate perception" (6, p. 275). The kind of situation set for *S*s apparently influences the degree of constancy. Brunswik (2) used a quite natural situation in which *S*s were instructed to select from a number of cubical blocks the one which appeared equal in size to the standard. In this natural situation, *S*s showed considerable constancy. Thouless (7) has been interested in the distinction between object-orientation and critical-stimulus orientation and found that the former is associated with an increase in constancy, whereas the critical-stimulus attitude is used more by professional artists and results in less constancy. Thouless hypothesized that the engineer, because of his practical training, will come closer to the "real" aspects of objects and will thus show more "phenomenal regression" toward these "real" characteristics. The increase of constancy with increase in age has been studied,

---

* Department of Psychology, University of Kentucky, Lexington.

[1] The author is appreciative of the statistical assistance given by the Computing Center, University of Kentucky.

Child Develpm., 1961, 32, 277-286.

but the results are conflicting. Osgood (6) reports that Burzloff, in a natural situation, using both adults and children, found (a) a high level of constancy and (b) no significant differences according to age. Osgood states: "Considerable interest has attached to the age variable . . . this whole question is very much up in the air at this time" (6, p. 279). The influence of intelligence may be concomitant with increase in age.

The present study approximates the natural situation and object-oriented attitude in the task presented and in the instructions given to *S*s and includes both children and adults. However, it is not concerned with the usual variables found in constancy experiments, namely, distance and retinal shape and angle.

This study investigated (a) accuracy of judgments of same-shaped figures and (b) the bases or cues used to estimate equality of different-shaped figures, e.g., comparing the size of a square with a circle or comparing the size of a circle with a triangle. Comparing here means comparing in direct perception, not by measurement or computation. There are several possibilities. A person may base his judgment on estimated areas or the diameter of a circle may be equated with the side of an equilateral triangle. Still another possibility is that a figure may be compared with another by inscribing or circumscribing. Such specific possibilities are termed the predesignated figures and are included in the variable series.

SUBJECTS

Table 1 presents the grade levels, ages, and number of *S*s in each group. None of the *S*s in the Arts and Sciences group had had geometry; all the engineering students had had geometry and were in their second semester of mechanical drawing. Each group except the engineers included males

TABLE 1

SUBJECTS

	MEAN AGE			
Grade	*Years*	*Months*	*Range*	*N*
Kindergarten	6	1	± 3 months	15
Second	7	3	± 3 months	15
Fourth	9	5	± 4 months	15
Sixth	11	5	± 3 months	15
Eighth	14	3	± 2 months	15
Arts and Sciences students				15
Engineering students				15
All subjects				105

and females. The children attend the University School of the University of Kentucky; from the experience of the author in intelligence testing at this school, a conservative estimate of the mean Stanford-Binet IQ is 120. The adults, of course, are college students.

METHOD

*S*s were presented a standard figure (a triangle, a square, or a circle) and were asked to select, from a series varying in size, called the variable series, the figure most nearly equal to the size of the standard figure. Each *S* made a total of 135 judgments. Forty-five of these judgments were similar-figure trials; i.e., the standard figure and the variables were the same shape. The remaining 90 trials were different-figure trials (e.g., the standard figure might be a triangle to be compared with a variable series of circles).

Both the standard series and the variable series were constructed of lemon yellow, antique finish, 80 weight, Buckeye cover paper obtained from the Beckett Paper Company, Hamilton, Ohio.

The standard series consisted of five of each of the following forms: equilateral triangles 15 inches on a side, 10 inches on a side, 5 inches on a side; squares, 15 inches on a side, 10 inches on a side, 5 inches on a side; and circles, 15 inches in diameter, 10 inches in diameter, and 5 inches in diameter. Thus, there were three shapes and three sizes of each shape making nine different combinations. Five random orders of this 45-member

TABLE 2

SIZES OF THE VARIABLE SERIES

Triangles		*Squares*		*Circles*	
22.80*	12.00	20.00	9.87*	20.00	9.00
21.21*	11.55*	19.00	9.19*	19.00	8.66*
20.20*	11.00	18.00	9.00	18.00	8.00
20.00	10.00*	17.00	8.66*	17.31*	7.42*
19.00	9.00	16.00	8.00	16.92*	7.00
18.00	8.66*	15.00*	7.00*	16.00	6.00
17.32	8.00*	14.00	6.58*	15.00*	5.64*
17.00	7.60*	13.26*	6.00*	14.00	5.00*
16.33*	7.00*	13.00*	5.00*	13.00*	4.33*
16.00	6.73*	12.00	4.43*	12.00	4.00*
15.00*	6.00	11.31*	4.00	11.55*	3.00*
14.00*	5.77*	11.00	3.54*	11.28*	
13.47*	5.00*	10.61*	3.00*	11.00*	
13.00*	4.33*	10.00*		10.00*	

NOTE.—Values indicate length in inches of sides of equilateral triangles, sides of squares, and diameters of circles.

* Predesignated values.

standard series were prepared; each *S* was examined with three of these orders.

There were three variable series: squares, circles, and triangles. The size gradations in linear measurements are shown in Table 2. The size intervals are not equal.

In each series, certain predesignated sizes are indicated by asterisks in Table 2. The other values in Table 2 were inserted to provide series of sufficient length and of a higher level of difficulty than was possible with only the predesignated sizes. Table 3 shows the precise values for figures of different shapes when different cues are used as the standard of equality;

TABLE 3

EXPLANATION OF PREDESIGNATED VALUES

Standard	*Variable*	*Possible Cues*	*15-Inch Size*	*10-Inch Size*	*5-Inch Size*
Triangle ..	Squares ...	Side of T = side of S	15.00	10.00	5.00
		Altitude of T = side of S	12.99	8.66	4.33
		Side of T = diagonal of S	10.61	7.07	3.54
		Altitude of T = diagonal of S ..	9.19	6.12	3.06
		Area of T = area of S	9.87	6.58	3.29
Triangle ..	Circles ...	Side of T = diameter of C	15.00	10.00	5.00
		Altitude of T = diameter of C ..	12.99	8.66	4.33
		Area of T = area of C	11.14	7.42	3.72
		C inscribed in T	8.66	5.77	2.89
		C circumscribed around T	17.31	11.55	5.77
Square ...	Triangles ..	Side of S = side of T	15.00	10.00	5.00
		Diagonal of S = side of T	21.21	14.14	7.07
		Side of S = altitude of T	17.32	11.55	5.77
		Diagonal of S = altitude of T ..	24.49	16.33	8.16
		Area of S = area of T	22.80	15.19	7.60
Square ...	Circles ...	Side of S = diameter of C	15.00	10.00	5.00
		Diagonal of S = diameter of C ..	21.21	14.14	7.07
		Area of S = area of C	16.92	11.28	5.64
Circle	Triangles ..	Diameter of C = side of T	15.00	10.00	5.00
		Diameter of C = altitude of T ..	17.32	11.55	5.77
		Area of C = area of T	20.20	13.47	6.73
		T inscribed in C	12.99	8.66	4.33
		T circumscribed around C	25.98	17.32	8.66
Circle	Squares ...	Diameter of C = side of S	15.00	10.00	5.00
		Diameter of C = diagonal of S ..	10.61	7.07	3.54
		Area of C = area of S	13.26	8.86	4.43

NOTE.—T = triangle, S = square, C = circle.

these are the predesignated values in Table 2. The predesignated figures were constructed to the nearest 0.1 inch; there was no figure constructed whose linear dimension was larger than 22.80 inches because material that large could not be obtained.

PROCEDURE

Each *S* was used for three sessions, spaced at least 24 hours apart. The sessions differed in the figures used as variables; in one session, triangles were used, in another, squares, and in the remaining session, circles were the variable figures. To control practice effects, each set of variable figures was used for the first, second, and third sessions by an equal number of *S*s.

In each session, 45 standard figures, in one of the random orders, were compared, one at a time, against a variable series. The variable series was arranged in order of size on a large table in front of *S*. On each trial the standard figure was placed on the table and *S*, who walked around the table at will, was told: "Look over all these figures and find the one which is the same size as this one." At the higher age levels, an occasional *S* asked what was meant by size. *E*'s reply was: "I want to find out what *you* mean. Please choose from the table the one which is the same size as this one."

The total number of judgments made was 14,175: *N* equals 105; there were five judgments per combination and three sizes, three shapes, and three standard-variable combinations.

RESULTS

Each *S*'s score is a mean of five trials; for each group, the mean is for 15 *S*s.

Equality Judgments for Similar-Form Figures

The accuracy of equality judgments for similar-form figures was high at all age levels. The kindergartners were as accurate as the college students. Variability decreased (a) with increase in age and (b) with decrease in size of figure. Table 4 shows the results for similar-form judgments.

Equality Judgments for Different-Form Figures

In a comparison of figures of different shapes it is not meaningful to establish an arbitrary measure of accuracy as a criterion. Some other basis or cue must be used. Table 3 shows the possible ways by which figures of different shapes could be compared. That is, *S* might use area or equal sides or altitude equal to diameter. The *t* test was used to compare the mean of each group for every size-shape comparison with each of the predesignated values or possible cues. Table 5 shows the apparent cue used for each standard-variable comparison. The last column shows the number of groups, out of a total of seven age groups, who apparently used the cue; i.e., no

Table 4

RESULTS FOR SIMILAR-FORM JUDGMENTS IN LINEAR INCHES

			GROUPS							
	Standard	*Variable*	*Kinderg.*	*2nd Grade*	*4th Grade*	*6th Grade*	*8th Grade*	*Coll. Stud.*	*Engineers*	*All Groups*
Mean	15-Inch T	T	14.9	15.5	15.4	15.6	15.5	15.8	16.0	15.5
SD			1.1	.9	.6	.6	.5	.5	.5	.8
Mean	10-Inch T	T	10.3	10.4	10.2	10.2	10.3	10.4	10.6	10.3
SD			.7	.6	.5	.8	.3	.4	.5	.6
Mean	5-Inch T	T	4.9	5.1	5.1	5.1	5.0	5.1	5.1	5.1
SD			.5	.6	.1	.1	.1	.2	.0	.3
Mean	15-Inch S	S	15.4	16.8	14.8	15.2	15.1	15.1	14.2	15.2
SD			1.3	2.4	.9	.6	.5	.5	2.0	1.6
Mean	10-Inch S	S	9.8	11.3	10.2	10.3	10.2	10.2	10.4	10.3
SD			.7	1.5	.4	.3	.2	.4	.4	.8
Mean	5-Inch S	S	4.8	5.6	5.1	5.0	5.0	5.0	5.2	5.1
SD			.4	1.0	.2	.0	.0	.1	.3	.5
Mean	15-Inch C	C	15.2	14.9	14.5	14.8	14.8	15.0	15.5	15.0
SD			.8	.6	.4	.4	.5	.3	.6	.6
Mean	10-Inch C	C	10.1	10.3	9.9	10.2	10.2	10.3	10.5	10.2
SD			.8	.6	.3	.3	.3	.4	.5	.5
Mean	5-Inch C	C	4.9	5.2	5.0	5.0	5.0	5.1	5.1	5.0
SD			.4	.6	.1	.0	.0	.0	.5	.3

NOTE.—T = triangle, S = square, C = circle.

TABLE 5

RESULTS FOR DIFFERENT-FIGURE COMPARISONS IN LINEAR INCHES FOR THE SEVEN GROUPS

Standard	*Variable*	*Mean*	*SD*	*Apparent Cue Used*	*No. of Groups Using Cue*
15 T	S	12.57	1.71	altitude = side	7
10 T	S	8.62	1.13	altitude = side	5
5 T	S	4.24	.53	altitude = side	6
15 T	C	12.92	1.84	altitude = diameter	7
10 T	C	9.13	1.34	altitude = diameter	5
5 T	C	4.52	.69	altitude = diameter	4
15 S	T	19.84	1.86	diagonal = side	2
10 S	T	13.90	2.11	diagonal = side	5
5 S	T	7.12	1.14	diagonal = side	6
15 C	T	18.94	2.13	equal areas	5
10 C	T	12.99	1.88	equal areas	5
5 C	T	6.60	1.16	equal areas	7
15 C	S	14.84	1.73	diameter = side	6
10 C	S	9.96	1.11	diameter = side	6
5 C	S	4.86	.58	diameter = side	4
15 S	C	16.46	1.52	equal areas	6
10 S	C	11.45	1.01	equal areas	6
5 S	C	5.78	.61	equal areas	6

NOTE.—T = triangle, S = square, C = circle. The seven groups are the seven grade levels, as kindergarten, 2nd grade, 4th grade, etc.

statistical difference between the mean of each group and the actual dimension for the specific cue was shown by the *t* test.

Size of figures and grade level of *S*s are omitted in Table 6.[2] Size did not influence the cue used; when a 15-, 10-, or 5-inch triangle was compared with squares, *S*s consistently, over all three sizes, equated altitude of triangle with side of square. Age or grade of *S*s did not influence the cue used; kindergarten through college students used the same apparent cues. All conclusions based upon *t* tests were confirmed by analyses of variance. No sex differences were found. *S*s were very consistent within themselves in overestimation or in underestimation as measured by the coefficient of concordance.

---

[2] Results are presented by groups and sizes in Table A which has been deposited as Document number 6605 with the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C. Copies may be secured by citing the Document number and by remitting $1.25 for photoprints, or $1.25 for 35 mm. microfilm. Advance payment is required. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 6

COMBINED RESULTS FOR DIFFERENT-FIGURE COMPARISONS

Standard	*Variable*	*Apparent Cue Used*	*No. of Groups Using Cue**
Square	Circle	equal areas	18
Square	Triangle	diagonal = side	13
Circle	Square	diameter = side	16
Circle	Triangle	equal areas	17
Triangle	Square	altitude = side	18
Triangle	Circle	altitude = diameter	16

* Total $N = 21$ groups (seven grade levels and three sizes).

DISCUSSION

The results suggest that fairly accurate judgment of similar-shaped figures develops before six years of age.

The results also suggest that the cues used as the basis for judgments did not differ according to age of *S*s or size of figures, but rather according to the figures themselves. Area was used when the standard was a square and circles were the variable and when the standard was a circle and triangles were the variable figures. Altitude was used when the triangle was the standard; i.e., altitude was made equal to the side of a square or the diameter of circles. In the other two comparisons, the longest dimension was most commonly used; i.e., the diagonal of the square was made equal to the side of the triangle and the diameter of circles was equated with the side of squares. Thus, multiple cues were used, and these cues seem to have depended on which shape was the standard and which was the variable.

In general, variability decreased with increase in age. Except for the engineering students, there was a fairly sharp break between the second and fourth graders, the older children showing less variability in their judgments. The engineering students walked around the table, looked at the variable stimuli, and muttered formulas to themselves. Of course, they did not know the sizes of the figures, and, therefore, these formulas were of no help. Their attempts to go beyond the task presented (comparing the figures spread on the table) by doing formulas in their heads when they did not know the true sizes of the figures, may not have remembered the formulas correctly, and may have made miscalculations probably accounted for the greater error in their judgments.

An interesting finding was the greater variability in judgments of squares and least variability in judgments of circles. This may be related to a decrease in the number of angles, from squares to triangles to circles.

Graphs were drawn showing each *S*'s score on each of the 27 combinations of standards, sizes, and variables to investigate whether the means were obscuring real differences in behavior. There was the possibility that some *S*s, for example, might equate the side of a triangle with the side of a square and that other *S*s might make the side of a triangle equal to the diagonal of a square. If this occurred, the mean for the group might indicate that the basis for judgment was altitude of triangle equal to side of square. The graphs show that this did not occur. Most *S*s' scores fell close to the point indicated by the mean of the group which, in turn, did not differ significantly from the value to be expected if altitude of triangles were equated with side of squares. These graphs are not presented since they merely confirm previous statistical analysis.

The possibility that area is the commonly used basis for comparing the size of all figures was not borne out by the results of this study. Area was used in only two of the six comparisons of different-shaped figures. Further validation appears in the responses of *S*s. At the end of the three sessions, each *S* was asked: "Can you tell me how you decided which square was equal to a triangle, etc.?" They frequently mentioned height, sides, and diameter as determinants of equality. They mentioned area in the judgment of circles, and the apparent cues in Table 4 indicate that area was used in two of the major comparisons of circles. Some *S*s declared they "never thought of area" as a basis for their judgments.

SUMMARY

*S*s ranging in grade from kindergarten through college were presented with a standard series of figures and were asked to select from a variable series the figure which was most nearly equal in size to the standard. Forty-five of these judgments were similar-figure trials; 90 were different-figure trials.

On the similar-figure trials (a) *S*s were accurate in estimation of equality of size; (b) young children were as accurate as adults; (c) variability of group judgments decreased with increase in age; (d) variability of group judgments decreased with decrease in size of the standard presented.

On the different-figure trials (a) the basis for judgments differed according to whether figures were in the standard or variable series; (b) cues differed according to the shapes of the figures; (c) the same cues were used by all groups regardless of size of figures; (d) there were no age or sex differences; (e) in one-third of the comparisons, area was used in estimating equality of size.

REFERENCES

1. BORING, E. G. *Sensation and perception in the history of experimental psychology*. Appleton-Century-Crofts, 1942.

2. Brunswik, E. Thing constancy as measured by correlation coefficients. *Psychol. Rev.*, 1940, 47, 69-78.

3. Hicks, J. A., & Stewart, F. D. The learning of abstract concepts of size. *Child Develpm.*, 1930, 1, 195-203.

4. Long, L. Conceptual relationships in children: the concept of roundness. *J. genet. Psychol.*, 1940, 57, 289-315.

5. Long, L. Size discrimination in children. *Child Develpm.*, 1941, 12, 247-254.

6. Osgood, C. E. *Method and theory in experimental psychology.* Oxford Univer. Press, 1953.

7. Thouless, R. H. Individual differences in phenomenal regression. *Brit. J. Psychol.*, 1932, 22, 216-241.

8. Thrum, M. E. The development of concepts of magnitude. *Child Develpm.*, 1935, 6, 120-140.

9. Welch, L. The development of discrimination of form and area. *J. Psychol.*, 1939, 7, 37-54.

INADEQUATE MOTHERING AND DISTURBANCE IN THE NEONATAL PERIOD[1]

LYNN OURTH *and* KENNETH B. BROWN*

University of Missouri

Much speculative writing has been done about adequate mothering in infancy. Vague terms such as "tender loving care," "emotional warmth," and "affectionate attitude" pervade the literature. But to the writers' knowledge, little controlled research has been done on determining the significant variables involved. For example, in the Yale rooming-in project (10) infants were placed with their mothers shortly after birth and the effects of the continuous maternal contacts observed. However, the investigators do not specify the variables involved, nor could they designate what produced positive effects in some instances and negative effects in others. A few investigators (e.g., 3, 9, 12) have made inferences about adequate mothering on the basis of deprivation studies, but here too the variables are only crudely identified. The mother or her behavior is viewed as being a "psychological organizer" (3), reliever of "physical tensions" and gratifier of "cravings for movement and stimulation" (9), and a provider of "emotional supplies" (12).

On the basis of her pioneer observations, Ribble (11) drew more specific conclusions than many more recent investigators. For her, the infant has an intrinsic need for contact with a mothering person and requires ample amounts of auditory, tactile, and kinesthetic stimulations for adequate development. However, Ausubel (1) is a recent critic of the "plentiful fondling" view. He believes that the evidence has not indicated there are persisting effects of maternal deprivation prior to six months of age. For Ausubel, intensive mothering is unnecessary during this early infantile period; adequate development occurs under ". . . a minimal amount of stimulation, personal attention and individualized care . . ." (1, p. 122). However, Ausubel does not detail what constitutes the "minimal amount."

---

* Department of Psychology, University of Missouri, Columbia.

[1] Based on a master's thesis study conducted by the first author under the supervision of the second author. The aid of Dr. Walter R. Hepner, formerly Professor of Pediatrics, University of Missouri Medical Center, is gratefully acknowledged.

Child Develpm., 1961, 32, 287-295.

Thompson (14), on the other hand, tentatively agrees with Ribble's general thesis. After reporting the McGill stimulus deprivation research on puppies and reviewing other animal studies on early environmental restrictions, Thompson concludes that ". . . restriction of experience occurring early in life definitely retards normal development, [however] it is not in itself stressful, if stress is taken to refer to some sharp change of the stimulus situation . . ." (14, p. 133).

Thompson further states, "Any radical change in a fairly well-structured environment will usually produce a strong affective reaction" (14, p. 133). In our view minimal stimulation of the human neonate probably constitutes a radical change from the continuous mild to moderate stimulations we believe are normally provided in the fetal environment. The fetus obviously is "wrapped in the firm support of the amnion." And as the mother breathes, walks, and otherwise changes position, even in sleep, the fetus is rather continuously subjected to mild and varied but more or less rhythmical cutaneous, kinesthetic, vestibular, and (to some extent) auditory stimulations. However, at present the evidence is quite tenuous that a need for such stimulations is acquired in this manner.

According to Carmichael (5), the fetus is capable of receiving cutaneous and kinesthetic stimulations by the twentieth week. Gesell and Amatruda's (8) findings suggest that the kinesthetic, cutaneous, and vestibular senses may be the only ones capable of responding to mild stimulations at birth. It appears possible that the cushioned, constantly varying pressure and positional changes imposed on the fetus play significant roles in the acquisition of the sensitivities present at birth, as well as in the acquisition of rudimentary needs for considerable stimulation of the types mentioned.

In summary, we may state that, although the long range consequences of mild stimulus deprivation as opposed to severe deprivation in early infancy still may be debatable, increased "affective reaction" or disturbed behavior should be an immediate effect of mild deprivation, if the above observations are correct. Since birth itself is an abrupt environmental change, it would appear that adjustment in the neonatal period should be facilitated by continuing the kinds and amounts of stimulations present during the fetal period, insofar as possible.

PROBLEM

The present experiment is concerned with the immediate effects of moderately different treatments of neonates during their brief hospital stay. One group was given no more handling than necessary in routine care and feeding, and the other was provided a moderate amount of "Ribble-type" mothering (moderate in terms of the amount assumed to be normally operating prior to birth). The experimental hypothesis was that the "non-mothered" (N-Mo) group would show more disturbed behavior than the "mothered" (Mo) group.

METHOD

Subjects

The *S*s were 20 neonates in the University of Missouri Medical Center's newborn nursery ward, divided equally into the Mo and the N-Mo groups. All were full term and biologically normal as determined by the staff pediatricians. Fourteen neonates were not used because they failed to meet these criteria; two were discarded at the age of two days because they acquired a respiratory disorder; and a few other prospective *S*s were not used because the *E* limited himself to a maximum of three *S*s at a time.

The *S*s were assigned to one or the other of the two groups alternately as they were admitted to the ward, insofar as this was possible consistent with placing prospective breast-fed infants in the Mo groups and prospective bottle-fed infants in the N-Mo group.

The decision on bottle- or breast-feeding was made jointly by the pediatrician in charge of each case and the mother. It was based on the mother's expressed preference and the pediatrician's judgment of her biological ability to nurse her infant. The mothers of all Mo group infants preferred breast feeding and were judged biologically capable on the basis of previous breast feeding, adequacy of nipple, etc. In the N-Mo group five of the mothers were judged biologically incapable of breast feeding and had consented to bottle feeding. The five remaining mothers preferred bottle feeding although judged biologically capable of breast feeding. Parenthetically, it is to be noted that all Mo infants received supplemental bottle feedings of Karo water for most of their hospital stay, i.e., nearly three days. Previous research (4, 6, 10, 13) indicates that type of feeding and feeding preference of the mother have no relation to the type of dependent variables investigated in the present study.

Only a secondary attempt was made at equating the groups for sex and skin color because previous research (2, 4, 7) suggests that these factors as such have no significant influence on the variables investigated here. In addition, biological normality and control of mode of feeding were considered more important for the study. The Mo group consisted of six males and four females, and the N-Mo group had three males and seven females. Three of the Mo group were white and seven brown, while the skin colors were equally divided in the N-Mo group.

Equipment and Materials

Nursery. The nursery was a soundproof room with dehumidified air and a constant temperature of 72° F. The 12 bassinets in the nursery were identical, consisting of transparent baskets mounted on stainless steel cabinets. A rocking chair and all the equipment needed for infant care were located in the room.

Materials. Medical record forms completed by the hospital staff provided information on age of the mother, ordinal position of the neonate, length of labor, initial weight, etc.

A behavior data sheet, constructed for the study, was used for determining the following data:

a. *Amount of crying,* in average minutes per hour, divided into subcategories according to the period when it occurred or to the apparent instigator. *Prefeeding crying* (stimulus unknown) was any crying, often accompanied by rooting and sucking, within 30 minutes of feeding time when no external stimulus could be found. *Postfeeding crying* (stimulus unknown) was that which occurred within 30 minutes after feeding and was not due to any determinable external stimulus. *Between feedings crying* (stimulus unknown) was crying which occurred for no determinable reason between the postfeeding and prefeeding periods. *External stimulus crying* was any crying associated with wet or soiled diapers, bathing, changing clothes, circumcision of males, picking up when asleep, and any other sudden or intense stimulation.

b. *Amount of sleeping,* in average minutes per hour. An infant was considered asleep when its eyes were closed, and it appeared relaxed and unresponsive to external stimulation.

c. *Amount of vomiting,* in average number of vomits per hour.

d. *Amount of handling,* in average minutes per hour. Total handling included nursing care and mothering care, the latter for the Mo group only.

Procedure

Treatment of nonmothered group. The treatment of the N-Mo group followed the usual hospital procedures, except for a few modifications to be noted later. Each infant was brought to the nursery within one hour after birth. The nurse on duty then administered initial nursing care, consisting of washing the infant's hair, putting saline solution in its eyes and hydrogen peroxide to its cord, and dressing it in a diaper and a long-sleeved cotton undershirt. Following this, she placed the infant on its abdomen in its bassinet and covered it loosely with two blankets. After 12 hours it was placed on an approximate four-hour feeding schedule, with the actual intervening time between feedings varying from 3.5 to 4.5 hours. At about 8:00 A.M. and 5:00 P.M. daily the ward physicians made rounds in the nursery. At one or the other of these rounds each infant was undressed and examined by a physician. Diaper changes and eye and cord care were performed whenever it was considered necessary. After the first day, each infant was bathed once daily.

During the fourth postnatal day, the male infants were circumcized. This took approximately one hour to perform. Because of the larger number of male infants in the Mo group, any negative effects of the circumcisions would bias the results against the experimental hypothesis.

The modifications from the normal hospital routine were the following: All N-Mo *Ss* were bottle fed on a standard hospital formula until they stopped sucking. Normally this took between 15 and 20 minutes. Bottle feeding was done rather than breast feeding in order to keep handling to a minimum. Originally it was planned to do all the feeding in the bassinet, but we had to compromise with current practice and let the mothers feed their infants for four of the six meals per day. Only the 2:00 A.M. and 6:00 P.M. bottles were given in the bassinets.

Each time the infant cried, either the *E* or the ward nurse on duty made a systematic check of the source, which was eliminated if it was an external stimulus. If the cause was determined to be hunger and the crying occurred within 30 minutes of the scheduled feeding time, the infant was fed. Otherwise the infant was allowed to cry, and the crying time was recorded. The nurses and aides were instructed to handle the infants of both groups no more than necessary in their nursing care.

Treatment of the mothered group. Treatment of the Mo group was similar to that for the N-Mo group in all respects except for breast feeding (with supplemental bottle feeding of Karo water for about the first three days) and the special handling given at feeding time. Since the mothers of this group were to provide the mothering handling of their infants when they were capable, careful instruction and practice was given them in the appropriate method. The *E* showed each mother how to hold the infant with its limbs in fetal position, firmly wrapped with blankets, and in close contact with her body. She also was instructed to rock it gently and rhythmically when it was through feeding and to avoid giving it any sudden stimulation. The *E* and the nurses made regular observations to make certain the mothers were providing handling that met the criteria of firm support, fetal positioning, and mild rhythmic stimulation. The nurses and aides were given similar instructions by the *E* in how to give mothering care to the Mo infants when it was necessary for them to take over from the mothers after feeding.

The total time allowed for feeding and mothering care of the Mo *S*s was 30 minutes at the 2:00 A.M. and 6:00 A.M. feedings and 60 minutes at the other four sessions. If some interference prevented the mother from handling her infant the full allotted time, the *E*, a nurse, or an aide completed the mothering care in a rocking chair upon the infant's return to the nursery, even if it was asleep. At other times the Mo infants received the minimal amount of handling necessary in nursing care.

E observed the *S*s of both groups throughout the study and made the appropriate entries on the data sheet except for the periods from 1:00 A.M. to 6:00 A.M., and at his mealtimes each day. During these times the nurse on duty made the observations and recordings, in accordance with instructions given her by the *E*.

RESULTS AND DISCUSSION

The variables considered important with regard to the equivalence of the Mo and N-Mo groups at birth were those indicative of biological health and reactivity. The measures used for determining initial equivalence were birth weight, birth temperature, number of reflexes at birth, length of labor, mother's age, and her number of children.

Because these data are only indirectly relevant to the study, it seems sufficient to state that *t* test analyses of the various measures on the two groups showed no difference which approached the .05 level. Insignificant initial differences were to be expected especially on variables of the *S*s (weight, temperature, and reflexes) because only full-term and biologically normal infants were used.

The results of direct relevance to the experimental prediction of more disturbance in the N-Mo group are those concerned with the treatments and reactions of the two groups during their stay in the hospital. As Table 1 shows, the infants of both groups were under the conditions of the study an average of only about 4.5 days. Congruent with the experimental design, the Mo group was handled an average of over three times as much as the N-Mo group (11.02 vs. 3.15 min. per hr.).

TABLE 1

COMPARISON OF THE NONMOTHERED AND MOTHERED NEONATES
(N = 10, each group)

	NONMOTHERED		MOTHERED			
Variable	*M*	*SD*	*M*	*SD*	*t*	*p*
Age at end of study (in days)	4.33	.65	4.61	.45	.99	
Handling (av. min./hr.)	3.15	.56	11.02	.84	24.51	.01
Crying (av. min./hr.)	5.42	.56	3.37	.32	9.51	.01
Prefeeding, unknown stimulus	1.95	.64	1.23	.53	2.72	.05
Postfeeding, unknown stimulus	.42	.34	.27	.18	1.18	
Between feedings, unknown stimulus	2.34	.33	1.08	.29	9.13	.01
External stimulus	1.31	.40	1.18	.48	1.31	
Vomits (av. no./hr.)	.02	.01	.01	.01	.66	
Sleep (av. min./hr.)	43.47	1.80	41.83	2.81	1.84	

The crying reactions of the two groups appear to support quite well the hypothesis of the study. As to total crying, the N-Mo infants cried an average of 5.42 min. per hr., and the Mo infants an average of 3.37 min. per hr. As shown in Table 1, the difference is significant at well beyond the .01 level ($t = 9.51$; t of 2.88 required for .01 level). Moreover, only one Mo infant averaged more crying than the lowest N-Mo infant (4.86 to 4.83 min. per hr.).

When the crying reactions for unknown reasons were isolated in terms of the period in which they occurred, it was found that the N-Mo group showed significantly more prefeeding crying (.05 level) and over twice as much crying between feedings ($t = 9.13$, significant at far beyond

.01 level). However, as Table 1 shows, the relatively small amounts of crying in the half-hour postfeeding periods were not significantly different.

The fact that both groups showed relatively little crying in the postfeeding period suggests that the feeding itself has an immediate tranquillizing effect which tends to mask any similar effect of the mothering given during and after feeding. The disturbing effect of insufficient mild, rhythmical stimulations at feeding time tended to be manifested later in the feeding cycle. Moreover, the insignificant group difference in crying associated with external "annoyances" (usually nursing care in the between feedings period) suggests that the different total amounts of mild stimulation received by the two groups has no direct relation to intervening reactions to sudden or intense stimuli.

No prediction was made concerning the relative amounts of sleeping expected of the two groups. This was because the relation between mild, rhythmical stimulation and sleep appeared complex and little known. For example, we might have predicted greater sleep in the Mo group because less disturbance was anticipated or less sleep because the extra stimulation would keep this group awake longer or less sleep because less fatigue was expected in this group. Whatever opposing factors may have been operating, if any, the results show that the Mo group averaged slightly less sleep, but the difference is not significant at the .05 level.

We had thought that the N-Mo group might show significantly more vomiting, but only if the lesser handling had been extremely disturbing (13). The fact that the groups did not differ significantly in vomiting suggests, as does the only moderate total amount of crying in the N-Mo group, that "routine" hospital treatment had some but not extremely disturbing effects on the immediate adjustment of neonates.

We may add parenthetically that the nonsignificant differences in both vomiting and postfeeding crying suggest the formula-by-bottle feeding was as "satisfying" to the N-Mo infants as the breast feeding was to the Mo group. This is in agreement with previous findings (e.g., 6, 13).

In general, we may conclude that the results of this study are consistent with a stimulus deprivation hypothesis which states that disruption of "familiar" stimulus patterns results in generalized tension or disturbed behavior. In this instance we were concerned only with a moderate lowering of the amounts of mild pressure, positional, and rhythmically changing cutaneous stimulations in the neonatal period from those which presumably impinge on the average fetus. Although the differences in disturbance found are far from extreme, the findings appear to be in the same direction as those of previous studies concerned with more severe and prolonged deprivation (e.g., 3, 7, 12).

In line with the above, the results also may be interpreted as being consistent with a prenatal primitive learning hypothesis concerning the development of sensory "needs." However, they of course do not directly support such an hypothesis any more than an "innate need" hypothesis.

More definitive research would be concerned with investigating directly the relation of variations in prenatal stimulation and amounts of mothering administered postnatally. For example, according to the prenatal learning hypothesis, infants who have had high amounts of prenatal stimulation should show more disturbance under low amounts of postnatal mothering than under high amounts of postnatal mothering; but the reverse difference probably should occur among infants who have been subjected to low prenatal stimulation. However, the problem of controlling nutritional and other possible extraneous variables between the two prenatal groups is an extremely difficult one; e.g., active mothers probably provide somewhat different nutriments than inactive mothers.

SUMMARY

An experiment was conducted to investigate certain characteristics of mothering and their effects on the human neonate. It was proposed that adequate mothering of newborns consists of continuing the patterns of stimulation which normally occur intrauterinally, especially mild pressure and rhythmically changing cutaneous stimulation. On this basis the hypothesis was proposed that a deficiency in such stimulation of the neonate results in disturbed behavior.

The *Ss* were two groups of 10 biologically normal neonates who were under the experimental conditions for about 4.5 days following birth. One group ("nonmothered") was handled no more than necessary in routine hospital care, and the other ("mothered") was treated the same way except for the administration of specified lengths of mild, firm support and rhythmical body stimulations at feeding times. The results showed that the nonmothered group cried significantly more than the mothered group, particularly when no external instigation was present. It was concluded that the results support the hypothesis; however, more definitive research will be necessary to determine whether differences in prenatal stimulus conditions are related to the type and amount of postnatal mothering needs.

REFERENCES

1. AUSUBEL, D. P. *Theory and problems of child development.* Grune & Stratton, 1957.
2. BAYLEY, N. A study of the crying of infants during mental and physical tests. *J. genet. Psychol.*, 1932, 40, 306-329.
3. BOWLBY, J. Maternal care and mental health. *World Health Organ. Monogr.*, No. 2, 1951.
4. BRYAN, E. S. Variations in the responses of infants during the first ten days of post-natal life. *Child Develpm.*, 1930, 1, 56-77.

5. CARMICHAEL, L. The onset and early development of behavior. In L. Carmichael (Ed.), *Manual of child psychology.* (2nd Ed.) Wiley, 1954. Pp. 60-185.

6. DAVIS, H., SEARS, R., MILLER, H., & BRODBECK, A. Effects of cup, bottle and breast feeding on oral activities of newborn infants. *Pediatrics,* 1948, 2, 549-558.

7. GATEWOOD, M., & WEISS, A. P. Race and sex differences in newborn infants. *J. genet. Psychol.,* 1930, 38, 31-49.

8. GESELL, A., & AMATRUDA, C. S. *Developmental diagnosis.* (2nd Ed.) Hoeber, 1947.

9. GOLDFARB, W. Emotional and intellectual consequences of psychologic deprivation in infancy: a revaluation. In P. H. Hoch & J. Zubin (Eds.), *Psychopathology of childhood.* Grune & Stratton, 1955. Pp. 105-119.

10. JACKSON, E. B., & KLATSKIN, J. Rooming-in research project: development of methodology of parent-child relationship study in a clinical setting. *Psychoanal. Study Child,* 1950, 5, 236-274.

11. RIBBLE, M. R. Infantile experience in relation to personality development. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. 2. Ronald Press, 1944. Pp. 621-651.

12. SPITZ, R. A. Infantile depression and the general adaptation syndrome. In P. H. Hoch & J. Zubin (Eds.), *Depression.* Grune & Stratton, 1954. Pp. 93-108.

13. STEWART, A. H., *et al.* Excessive infant crying (colic) in relation to parent behavior. *Amer. J. Psychiat.,* 1954, 110, 687-694.

14. THOMPSON, W. R. Early environment—its importance for later behavior. In P. H. Hoch & J. Zubin (Eds.), *Psychopathology of childhood.* Grune & Stratton, 1955. Pp. 120-139.

ASSOCIATIVE CHARACTERISTICS OF SIXTY-THREE ADJECTIVES AND THEIR RELATION TO VERBAL PAIRED-ASSOCIATE LEARNING IN CHILDREN[1]

ALFRED CASTANEDA,* LEILA SNYDER FAHEL,
and RICHARD ODOM[2]
University of Texas

Considerable information regarding verbal learning phenomena has been derived from studies employing the paired-associate method of learning. Fruitful use of this method has been made in conjunction with adult (college age) populations in the empirical and theoretical analyses of such areas of investigation as learning, motivation, and personality. Although it would seem readily adaptable, and considering that the only major requirement determining selection of subjects to be studied by this method is the ability to read, it is surprising that it has not received systematic attention in connection with younger (elementary school age) populations. Characteristically, work with this method has employed verbal material (word lists), e.g., adjectives, nonsense syllables, etc., whose associative characteristics have been assessed. Examples of word lists which have been standardized on adult populations and used extensively in paired-associate as well as serial learning experiments are those described by Hilgard (1). However, comparable lists which have been standardized on younger populations are not currently available.

The present report describes the associative characteristics for 63 common adjectives obtained from a population of fourth, fifth, and sixth grade children. The procedure that was used is an adaptation from one followed by Russell and Storms (2) in their development of norms of associative strength in a college age population. Basically, the procedure represents a variation of the word-association method. The frequencies with which given associations occurred to each adjective are described. In order to assess the utility of the present list for paired-associate learning in children, two lists consisting either of high or low association values between the

---

* Department of Psychology, University of Texas, Austin.

[1] This study was supported by grants from the Hogg Foundation, University of Texas, and the National Institute of Mental Health (M-4240).

[2] The writers wish to express their appreciation to Mr. T. Houston Foster, Principal, I. W. Popham School, Del Valle, Texas, and Sister Julia, Principal, St. Austin's School, Austin, Texas, for their assistance and cooperation.

Child Develpm., 1961, 32, 297-304.

word pairs were employed in a paired-associate experiment involving fifth and sixth grade children.

METHOD

The experimental list was composed of a total of 69 common adjectives judged to be within the range of reading ability of fourth, fifth, and sixth grade children. However, six adjectives are omitted from the present report since approximately 13 to 22 per cent of the *S*s in the present sample failed to read these words correctly.

The words were typed in upper case letters and exposed by means of a Hunter Card Master (Model 340). Each word was exposed for 8 sec. with 2 sec. between exposures and a 10-sec. interval between exposure of the 23rd and 24th words and between the 46th and 47th words. Three different orders of exposing the words were used, and *S*s were randomly assigned to one of these conditions.

The *S* was required to call out any words that the adjective made him think of after first having read it aloud. Each *S* was given a practice list in order to insure that he comprehended what was to be required of him and to identify those *S*s with marked reading disabilities. During the practice list *S* was prompted to give single words rather than phrases and encouraged to give as many word-associations as possible within the 8-sec. interval. The practice list consisted of 10 words exposed in the following order: CAT, BOY, RUN, DOG, GIRL, ONE, SPOT, BROWN, JUMP, MOTHER. All *S*s met a criterion of nine or more words read correctly and one association given to each word on the practice list.

A total of 95 *S*s participated. However, the data of 13 *S*s from the total sample were excluded. These *S*s either failed to read correctly approximately 50 per cent of the first 23 words on the experimental list or failed to give associations to approximately 50 per cent of the first 23 words.

RESULTS

The 63 adjectives are listed alphabetically. Following each adjective are the associations given to it in order of decreasing frequency. Frequency is given as the percentage (in parentheses) of the *S*s from the final sample of 82 who gave the word as an association. In order to facilitate use of the present list, associations which also appear as stimulus items in the list are reproduced in capital letters. Also, an association which is asterisked indicates that it occurred as an association to one or more of the other stimulus items in the list. All associations which were either proper nouns or pronouns have been excluded. Finally, all associations which occurred with a frequency of one (1.2 per cent) are not reported. Since there were over 2,000 such associations they are omitted for reasons of space.[3]

---

[3] Copies of these are available upon request.

Stimulus	*Response*
ABLE	do(12) can(11) unable(8) can't(2) good*(2) well(2)
AFRAID	scared*(61) frightened(16) brave(8) fear(4) unafraid(4) coward(2) ghost(2) good*(2) happy*(2) knife(2) nervous(2)
ANGRY	mad*(61) happy*(12) mean*(6) BAD*(5) SAD*(5) good*(4) NICE*(4) cruel(2) glad*(2) unhappy*(2)
BAD	good*(37) mean*(18) naughty(11) NICE*(6) ANGRY(4) WRONG(4) COLD*(2) mad*(2) terrible(2)
BEAUTIFUL	PRETTY*(62) ugly*(23) NICE*(10) lovely*(7) beauty*(6) wonderful*(6) flower*(5) girl*(4) gorgeous*(4) AWFUL*(2) butterfly(2) house*(2) tall*(2)
BEST	good*(46) worst(8) BAD*(7) NICE*(7) clothes*(4) AWFUL*(2) BETTER(2) FINE*(2) flower*(2) friend*(2) less*(2) like*(2) wonderful*(2)
BETTER	good*(24) worse(13) BAD*(5) NICE*(5) AWFUL*(4) well*(4) work(4) BEST*(2) butter(2) FAIR*(2) FINE*(2) OLD*(2)
BIG	LARGE*(37) LITTLE*(28) SMALL*(24) tall*(23) huge*(8) medium*(7) FAT*(6) enormous*(5) giant*(5) round*(5) high*(4) gigantic*(2) man*(2) STRONG(2)
BRIGHT	LIGHT*(33) sun*(17) moon*(16) DARK*(15) shiny*(7) PRETTY*(6) sky*(6) dull(5) star(5) day*(5) color*(4) dim*(4) night*(4) black*(2) colorful(2) glaring(2) NEW*(2) smart(2) sunny*(2)
BROKEN	fixed(12) cracked(8) glass*(7) leg(7) break(6) broke(6) arm*(5) HURT(5) dish(4) unbroken(4) vase(4) car*(2) dropped(2) fell(2) half(2) pieces(2) shatter(2) toys*(2)
CLEAN	dirty*(43) filthy(5) CLEAR(4) clothes*(4) dirt(4) DRY*(4) bath*(2) house*(2) mean*(2) NICE*(2) PRETTY*(2) room*(2) sanitary(2) wash*(2) white*(2)
CLEAR	glass*(10) water*(10) BRIGHT*(7) CLEAN*(7) DARK*(5) see(5) sky*(5) cloudy(4) dirty*(4) NICE*(4) blue(2) clouds(2) COOL*(2) day*(2) FAIR*(2) foggy(2) LIGHT*(2) sunny*(2) WARM*(2) wash*(2) white*(2) window(2)
COLD	HOT*(33) WARM*(28) freezing*(23) COOL*(21) snow(7) ice(6) winter*(6) chilly(5) freeze(5) froze(4) weather*(4) rain*(2) shiver(2) water*(2) WET*(2)
COOL	COLD*(50) HOT*(37) WARM*(24) freezing*(7) NICE*(4) water*(4) winter*(4) air(2) aircondition(2) boiling*(2) fresh(2) freezer(2) spring*(2) weather*(2)
DARK	night*(38) LIGHT*(37) black*(15) moon*(8) day*(6) BRIGHT*(5) brown(2) CLEAN*(2) green(2) room*(2) scared*(2) stars*(2) sun*(2) sunny*(2)
DEEP	shallow(22) hole(18) water*(15) down*(7) LONG*(6) high*(5) ground(4) BIG*(2) DARK*(2) far*(2) ocean*(2) sea(2) SHORT*(2) steep(2) up*(2) WIDE*(2)
DIFFERENT	SAME*(22) alike*(8) people*(5) STRANGE(5) something*(4) unusual*(4) colors(2) girls(2) kind*(2) NEW*(2) other*(2) things*(2)

Stimulus	*Response*
DRY	WET*(44) damp*(7) clothes*(6) water*(6), CLEAN*(5) desert(4) HOT*(4) WARM*(4) land(2) soak(2) sun*(2) sunny*(2) weather*(2)
ENOUGH	much*(7) plenty*(7) lots*(6) all(5) MORE*(5) food*(4) FULL*(4) books*(2) eggs(2) had(2) HUNGRY(2) less*(2) not(2) through(2)
EVEN	STRAIGHT*(13) uneven(7) crooked(6) numbers*(5) good*(4) LONG*(4) SAME*(4) fine(2) LARGE*(2) level(2) odd*(2)
FAIR	good*(10) carnival(8) weather*(7) circus(6) cheat(5) fun*(5) game(5) right*(5) honest(4) NICE*(4) animal*(2) chair(2) fairy(2) mean*(2) play*(2) unfair(2)
FAST	SLOW(41) run*(38) running*(11) go*(6) jump(6) play*(5) speed(4) stop*(4) hurry(2) skip(2) swift*(2) quickly(2)
FAT	skinny*(34) BIG*(22) chubby(12) round*(12) medium*(10) LARGE*(7) SMALL*(6) thin*(6) pig(5) LITTLE*(4) boy*(2) dogs(2) man*(2) SHORT*(2)
FEW	MANY*(21) lots*(16) LITTLE*(13) MORE*(8) less*(7) much*(4) several(4) three(4) hundred(2) LEAST*(2) one*(2) SMALL*(2) things*(2)
FINE	good*(24) NICE*(7) well*(7) BAD*(5) PRETTY*(4) SOFT(4) BEST*(2) coarse(2) great(2) okay(2) mine(2) something*(2) thin*(2)
FREE	freedom(4) jail(4) SAFE(4) ABLE(2) caught(2) cost(2) happy*(2) money*(2) OLD*(2) pay(2)
FULL	empty*(20) ENOUGH*(8) eat*(4) fill(4) filled(4) plenty*(4) water*(4) basket(2) FAT*(2) glass*(2) happy*(2) lots*(2) milk(2) overflow(2) top(2) up*(2)
FUNNY	laugh(17) silly(16) clown(11) SAD*(7) laughing(6) fun*(5) happy*(5) giggle(4) people*(4) gay*(2) party(2) smile(2) unhappy*(2)
HOT	COLD*(32) WARM*(28) COOL*(17) sun*(10) boiling*(8) stove(6) burn(5) DRY*(5) fire*(5) burning(4) day*(4) steam(4) steaming(4) sunny*(4) heat*(2) heater(2) sweat(2) water*(2) WET*(2)
HUNGRY	starve(21) food*(17) starving(11) full*(6) thirsty(6) eat*(5) animal*(2) empty*(2) POOR*(2)
HURT	cut(12) injured(10) BROKEN(6) BAD*(5) sore(5) cry(4) skin(4) arm*(2) AWFUL*(2) bleeding(2) crying*(2) feelings(2) good*(2) scratch(2) SICK*(2) unhurt(2)
LARGE	BIG*(56) SMALL*(26) tall*(13) huge*(8) enormous*(7) LITTLE*(7) medium*(7) gigantic*(5) giant*(4) high*(4) round*(4) building(2) FAT*(2)
LIGHT	DARK*(38) BRIGHT*(24) sun*(16) day*(10) moon*(10) night*(6) bulb(5) morning(4) darkness(2) dim*(2) HOT*(2) shiny*(2) stars*(2) white*(2) yellow(2)
LITTLE	SMALL*(60) BIG*(34) SHORT*(16) tiny*(12) tall*(10) LARGE*(8) baby*(7) medium*(6) FAT*(5) midget(2) people*(2) thin*(2) WIDE*(2)

Stimulus	*Response*
LONG	SHORT*(37) tall*(16) BIG*(12) STRAIGHT*(7) WIDE*(7) SMALL*(6) far*(5) LARGE*(5) mile(5) rope(5) skinny*(5) high*(4) thin*(4) FAT*(2) line(2) LOW(2) medium*(2) road*(2) slender(2) stick(2)
LOW	high*(37) SMALL*(20) BIG*(8) down*(7) LITTLE*(6) DEEP(5) SHORT*(5) medium*(4) tall*(4) LONG*(2) mountains(2) skinny*(2) under(2)
MANY	lots*(28) FEW*(18) much*(10) MORE*(6) things*(6) less*(5) people*(5) LITTLE*(4) LARGE*(2) money*(2) none*(2) numbers*(2) pencils(2) SMALL*(2) toys*(2)
MORE	lots*(16) less*(12) MANY*(10) ENOUGH*(5) LITTLE*(5) money*(5) plenty*(4) books*(2) candy*(2) empty*(2) FAST*(2) FEW*(2) for(2) people*(2) POOR*(2) want(2)
MOST	lots*(24) MORE*(16) MANY*(7) less*(10) LITTLE*(5) much*(5) FEW*(4) LEAST*(4) things*(4) almost(2) ENOUGH*(2) none*(2)
NEW	OLD*(41) PRETTY*(8) dress*(6) shoes(6) car*(4) shiny*(4) bought(2) boy*(2) BRIGHT*(2) girl*(2) good*(2) house*(2) NICE*(2) skirt(2)
NICE	good*(32) BAD*(16) kind*(13) SWEET*(10) mean*(8) PRETTY*(5) mice(4) polite(4) AWFUL*(2) dog*(2) happy*(2) helpful(2) lady*(2)
OLD	YOUNG(30) NEW*(20) grandfather(6) age*(4) grandmother(4) COLD*(2) elderly(2) feeble(2) gray(2) lady*(2) middleaged*(2) people*(2) torn(2) used(2)
POOR	RICH(48) money*(10) OLD*(5) wealthy*(4) house*(2) NICE*(2) SICK*(2)
PRETTY	BEAUTIFUL*(43) ugly*(24) NICE*(13) handsome(5) beauty*(4) cute(4) gorgeous*(4) lovely*(4) color*(2) dress*(2) FUNNY*(2) girl*(2) hair*(2) look(2) SWEET*(2)
QUIET	noisy(23) loud(10) STILL(10) library(4) people*(4) listening(2) peaceful(2) talk(2)
REAL	true(10) alive(8) fantastic(4) play*(4) book(2) cat*(2) dead(2) fables(2) fake(2) fiction(2) horse(2) live(2) money*(2) people*(2) right*(2) unbelievable(2)
RICH	POOR*(44) money*(23) gold(6) wealthy*(6) lots*(4) people*(4) happy*(2) house*(2) millionaire(2) plentiful(2) PRETTY*(2) silver(2)
SAD	happy*(33) unhappy*(23) glad*(13) good*(6) crying*(5) joyful(5) mad*(5) sorry(5) BAD*(4) gloomy(4) lonely(4) gay*(2)
SAFE	unsafe(7) money*(5) street(5) careful(4) danger(4) dangerous(4) policeman(4) safety(4) bank(2)
SAME	DIFFERENT*(12) alike*(11) like*(6) thing*(6) clothes*(5) dress*(2) one*(2) other*(2) person*(2) twins(2) unalike(2) unusual*(2)
SHORT	SMALL*(32) tall*(28) LITTLE*(22) LONG*(16) BIG*(11) FAT*(10) medium*(6) high*(4) skinny*(4) hair*(2) LARGE*(2)
SLOW	FAST*(50) car*(8) walk(7) go*(5) run*(5) medium*(4) pokey(4) stop*(4) down*(2) running*(2) swift*(2) turtle(2)

Stimulus	*Response*
Small	LITTLE*(49) BIG*(23) tiny*(13) LARGE*(10) tall*(10) baby*(8) SHORT*(7) child*(4) children(2) doll(2) FAT*(2) size(2) skinny*(2) thin*(2)
Soft	hard(32) fluffy(11) bed(10) pillow(8) smooth(6) feathers(5) cushion(4) kitten(4) rough(4) BEAUTIFUL*(2) furry(2) fuzzy(2) medium*(2) NICE*(2) ugly*(2)
Still	QUIET(21) move(11) moving(11) BAD*(2) FAST*(2) read(2) sitting(2) stand(2) water*(2) wiggle(2) working(2)
Strange	DIFFERENT*(8) known(6) friend*(5) stranger(5) man*(4) NEW*(4) queer(4) thing*(4) unusual*(4) feeling(2) FUNNY*(2) mysterious(2) odd*(2) person*(2)
Strong	weak(26) muscles(16) healthy(15) BIG*(11) man*(5) FAT*(4) heavy(4) powerful(4) good*(2) lift(2) LITTLE*(2) men(2) tough(2)
Sweet	NICE*(20) good*(17) sugar(13) sour(13) candy*(10) apple(7) bitter(5) orange(5) BAD*(4) PRETTY*(4) BEAUTIFUL*(2) flower*(2) gentle(2) kind*(2) mean*(2) salt(2) taste(2)
Warm	HOT*(57) COLD*(35) COOL*(12) sun*(7) heat*(5) boiling*(4) sunny*(4) blanket(2) coat(2) DRY*(2) fire*(2) spring*(2) summer(2) water*(2) weather*(2)
Wet	DRY*(37) rain*(12) water*(12) damp*(11) rainy(7) COLD*(5) sopping(5) swimming(4) bath*(2) CLEAN*(2) dirty*(2) floor(2) lightning(2) ocean*(2) SICK*(2) soaked(2) soaking(2) thunder(2)
Wide	BIG*(26) LARGE*(13) LONG*(8) thin*(7) broad(6) FAT*(6) narrow(5) skinny*(5) tall*(5) LITTLE*(4) room*(4) round*(4) SMALL*(4) ocean*(2) open(2) road*(2) width(2)
Wrong	right*(51) BAD*(7) answer(4) good*(4) AWFUL*(2) did(2) shouldn't(2) something*(2)
Young	OLD*(49) LITTLE*(7) age*(5) baby*(5) boy*(5) NEW*(5) child*(4) girl*(4) middleaged*(4) BIG*(2) cat*(2) dog*(2) NICE*(2) people*(2) SMALL*(2) tall*(2) twelve(2)

Relation to Paired-Associate Learning

Two sets of six word pairs (both of which contain common stimulus items) were constructed from the present list. One set consisted of word pairs of high association values, the other of low association values. These pairs are presented in Table 1 along with their association values in percentages.

The words were presented by means of the Hunter Card Master. A 3-sec. anticipation interval during which the stimulus item was exposed preceded a 3-sec. exposure of the response item. There was a 2-sec. interval between exposures of the pairs and 8 sec. between trials. A trial consisted of one exposure of each of the six pairs. In order to prevent the pairs from being learned serially, three different orders of exposing the pairs were employed. The *S*s were carried to a criterion of one perfect (errorless) trial or to a maximum of 15 trials.

TABLE 1

WORD PAIRS AND THEIR ASSOCIATION VALUES
(IN PARENTHESES)

	RESPONSE	
Stimulus	*Group I*	*Group II*
ANGRY	mad (61)	kind (1)
BEAUTIFUL	ugly (23)	sorry (1)
BIG	tall (23)	thin (1)
DARK	night (38)	far (1)
FAST	run (38)	high (1)
RICH	money (23)	good (1)

A total of 50 fifth and sixth grade *S*s were randomly assigned to one of two groups. Group I consisted of 28 *S*s and group II of 22. The list for group I consisted of word pairs of high association values and of low association values for group II.

A precautionary measure designed to identify those *S*s unable to read the words comprising the experimental list was employed. The appropriate words were typed in alphabetical order on separate sheets, and the *S* was required to read these words immediately prior to the paired-associate task (*see* instructions). All *S*s read the words without error.

The instructions to *S* were as follows: "This is a game to see how well you can remember. This is how it goes. Here on this side (*E* points to left aperture) a little window will open up, and you will see a word. A few seconds after this window opens, the window on this side (*E* points to right aperture) will open and you will see a *second* word. Your job is to guess what the second word is before its window opens. You say the word out loud so that I can hear you. (*E* hands *S* the words used in the experiment prepared in alphabetical order.) Here are the words that we will use. Would you read each one for me first. Fine, now we are ready to start. Remember, you are to say out loud the second word which is going to appear in the second window before the window opens. In the beginning you will make a lot of mistakes, but as you learn how it goes, you won't make so many mistakes."

RESULTS AND DISCUSSION

In terms of the mean number of trials to reach the criterion of one perfect recitation, the means for groups I and II were 3.4 and 11.0, respectively. The difference between the means was significant at the .001 level ($t =$ 10.76). These results indicate that performance in paired-associate learning was better where the association values between the stimulus-response items were initially high.

These results suggest that one of the parameters found to be important in work with adults, i.e., associative strength, may be varied in paired-associate learning in children by use of the present list or other lists constructed by comparable methods. The present list, however, is limited in the number of adjectives, and additional research concerned with increasing the number of adjectives in the present list, as well as constructing other lists appropriate in reading level for children in the lower grades, would be desirable. Furthermore, the adaptations necessary for work with this method in child populations of different ages would require research on those variables more immediately connected with the paired-associate method, e.g., anticipation intervals, rates of presentation, presence of competing associations, instructions, length of list, etc. Knowledge of the variables would be useful to studies more directly concerned with other factors, e.g., motivational and personality factors. As indicated earlier, these latter factors have been studied in adult populations with the paired-associate method, and there appear to be no *a priori* reasons why they could not be studied in conjunction with this method in younger populations.

SUMMARY

The present report describes the word associations and the relative frequency with which they occurred in a sample of 82 fourth, fifth, and sixth grade children to a list of 63 common adjectives. Also, two lists composed either of high or low association values between the to-be-learned stimulus and response items were constructed from the standardization list and employed in a paired-associate learning experiment with fifth and sixth grade children. Learning by the anticipation method was found to be significantly better the higher the association values.

REFERENCES

1. HILGARD, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. Wiley, 1951. Pp. 517-567.
2. RUSSELL, W. A., & STORMS, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.,* 1955, 49, 237-293.

VARIABLES IN EARLY DISCRIMINATION LEARNING: III. SIMULTANEOUS VS. SUCCESSIVE STIMULUS PRESENTATION[1,2]

WENDELL E. JEFFREY*

University of California, Los Angeles

In a previous study in this series it was found that children had great difficulty discriminating between tones of different pitch (4). It was shown that at least part of the difficulty was due to the response required of *S*. Another condition of particular concern for the present experiment was that of the method of stimulus presentation; the stimuli were presented successively, a response being required to each stimulus.

A number of studies have contrasted simultaneous and successive methods of stimulus presentation in the study of discrimination learning, and a variety of results have been obtained (1, 2, 5, 6). The various procedural differences, as well as stimulus and subject differences, make the interpretation of these results difficult. Inasmuch as Loess and Duncan (5) obtained a difference in learning speed between simultaneous and successive conditions with adult *S*s when the discrimination was a difficult one, it seemed likely that such a difference would also be found with children. If it could be established that children do find the successive presentation condition more difficult, this would provide the opportunity for studying some of the factors that might be responsible for the difference in difficulty.

The stimulus presentation conditions as used by Loess and Duncan were essentially as follows: when the stimuli were presented simultaneously, *S* was required to approach one or the other stimulus, regardless of its position to the right or left, in order to receive reinforcement. When the stimuli were presented successively, a single stimulus at a time, a different response was required for each stimulus. Thus, this latter condition not only provided less opportunity for comparison of the stimuli but also called for a different type of response from *S*.

Therefore, the present study was designed to test the hypothesis that discrimination learning with successive presentation of the stimuli is more difficult than when the stimuli are presented simultaneously and that this difficulty can be at least partially accounted for in terms of the response called for under the successive presentation condition.

---

* Department of Psychology, University of California, Los Angeles 24.

[1] The writer wishes to express his appreciation for the cooperation of Mr. Sterling Stott and Mrs. Margaret D. Mathews of the University Elementary School and Kindergarten.

[2] This research was supported by grant G-4414 from the National Science Foundation.

Child Develpm., 1961, 32, 305-310.

METHOD

Subjects. Twenty-six children from the Junior Kindergarten of the University Nursery School served as *Ss*. The children were randomly assigned to three groups with nine each in groups I and II and eight in group III. The mean ages of these groups were 5 years, 4 months; 5 years, 3 months; and 5 years, 4 months, respectively.

Apparatus. The stimuli were square blocks, one with an area of 2.5 sq. in. on a side, and the other with an area of 2.0 sq. in. on a side. An opaque screen shielded the stimuli from *S*'s view between trials. Two push buttons, mounted 14 in. apart on a board 2 in. wide and 16 in. long, could be placed in front of *S* with buttons to the right or left or one in front of the other. Programming apparatus turned on or continued tape recorded music, used as an incentive, if the correct button was pushed, while pressure on the incorrect button turned off the music. Once off, the music stayed off until a correct response was made on a subsequent presentation of the stimuli.

Procedure. Group I was shown the stimuli successively. A single stimulus was placed between the two buttons which were oriented to the right and left of *S*. *S*'s task was to learn to press the button on the left for one of the stimuli and the button on the right for the other. The correct response for each stimulus and the one presented first was counterbalanced among *Ss*. Prior to the training trials, *S* was shown the blocks separately and *E* pointed out the correct button for each block without labeling either the blocks or the buttons as indicated in the instructions that follow.

> I have two blocks here behind the screen. When I show you one of them, one button will make the music come on, but when I show you the other block, the other button will make the music come on. So, when I show you this block (a block is presented) you should press this button (a button is pointed to) to make the music come on, but (the first block is removed and the second presented) when I show you this block, this button should be pressed (*E* points to other button). Try to press the correct button each time so you will keep the music on. If you make a mistake the music will go off, but if you press the correct button the next time you can make the music come back on. Look carefully at the blocks so you will know which button to press.

The group II *Ss* were presented with two stimuli simultaneously but separated by 12 to 13 in. Thus, each block was approximately 1 in. from a button. Their task was to press the button beside the positive stimulus. For four *Ss* the positive stimulus was the small block, and for five *Ss* it was the large block. Prior to beginning the training trials, the blocks were placed in the positions indicated above, and the correct button was indicated with the positive stimulus both on the right and on the left; i.e., *Ss* saw the stimuli in both positions. As with group I, the instructions were given as follows without labeling either the button or the block.

> I have two blocks here behind the screen. I am going to place them in front of you, like this (the blocks are presented). In order to make the music

> come on, you must press the button near this block (*E* indicates the block) which would be this button (*E* indicates the button), not the button near this block (*E* indicates the negative block). So this way you would press this button (*E* indicates appropriately) but (*E* reverses the blocks) this way you would press this button (*E* indicates the other button). Try to get it right each time so the music will stay on.

Group III made it possible to evaluate the response factors involved in a successive discrimination and permitted some control of the factor of distance from the stimulus to the response device. Under the condition of simultaneous presentation in group II, *S* made only a simple approach response to the positive block, whereas in the successive presentation in group I, although the response required was physically the same, *S* had not only to discriminate the stimuli but also to learn whether a particular response went with one or the other stimulus. For group III there was only a single button in front of the screen. The stimuli were presented successively 5 to 6 in. behind the button, and *S* was given the following instructions.

> I have two blocks here behind the screen. When I show you this block you should press the button to make the music come on, but (the first block is removed and then the second is presented) when I show you this block you should not press the button or the music will go off. Try to get it right each time so the music will stay on.
>
> Look carefully at the blocks so you will know whether to press the button or not. If you make a mistake the music will go off, and you will have to wait for the next block to make the music come back on. Look carefully at the blocks so you will know whether to press the button or not.

The stimuli in all instances were presented at intervals of approximately 8 sec. following the order LRLLRLRR or + — + + — + — —. Training for all groups proceeded to a criterion of 10 successive correct responses. Immediately after reaching the criterion, the *S*s of each group were then given the following transfer tasks with appropriate instructions for each. Group I was given the original task of group II, and groups II and III were given the original task of group I.

RESULTS

Table 1 provides the median trials to criterion for both the training and transfer tasks for each group. Considering the trials on the first task alone, it should be noted that for group I five of nine *S*s refused to continue playing at anywhere from 14 to 61 trials, the median being 37 trials. This result has been obtained previously with similarly difficult tasks (3, 4). Of the rest of the *S*s in group I, one *S* made no errors following instructions and the other *S*s took 39, 40, and 72 trials to reach the criterion of 10 successive correct responses. The median for these latter four scores is given in Table 1. In contrast to the performance of the *S*s in group I, no group II *S* failed to learn the discrimination; the range of trials was from 0 to 20. Of the group

TABLE 1

MEDIAN TRIALS TO CRITERION UNDER EACH CONDITION FOR EACH GROUP

Group I			*Group II*		*Group III*	
Succ	∞	(39.5*)	Sim	3	Succ (r-nr)	7.5
Sim	∞	(10.0*)	Succ	2	Succ	6.0
Total	∞	(58.0*)	Total	9	Total	17.5

* These medians were calculated for the four *S*s who met the criterion. The analysis of variance included the *S*s who failed to reach the criterion.

III *S*s, one *S* refused to continue playing, but the rest learned the discrimination in from 0 to 31 trials. Thus, all *S*s in group II and all except one in group III reached the criterion in fewer trials than the median trial at which the five *S*s of group I refused to continue playing. A Kruskal-Wallis analysis of variance permitted the rejection of the hypothesis of no difference in the sums of rank scores with $p < .05$. The group II — group III difference is not significant.

Probably the most interesting and most rigorous comparison among groups can be made by using the total number of trials for each group to reach the criterion under the successive presentation condition (*see* Table 1). This means that for group II the trials to, but not including, the criterion with simultaneous presentation, and for group III the trials to criterion with the response-no response successive condition are also included in the total score. Those *S*s who failed to continue at any point were given a score of infinity. The Kruskal-Wallis analysis of these scores permits the rejection of the null hypothesis with $p < .05$. Further analysis of the differences among individual groups on trials to reach the criterion on the successive presentation task shows a clearly significant difference only between groups I and II ($p < .05$) when the Mann-Whitney *U* test is used. The difference between groups II and III produces a *U* with $.05 < p < .10$. All tests are two tailed.

It is also of interest to compare learning on the second task alone (*see* Table 1). It should be noted that the group I *S*s had a median of 10 trials to criterion with the simultaneous condition alone even after training on the successive condition; whereas group II *S*s had a median of only nine trials to criterion when trials to criterion were combined on both tasks. Thus, the simultaneous condition was not only an easier one on which to learn originally, but in this case the *S*s who learned the simultaneous task originally were able to learn both tasks faster than the group that learned the successive first could learn the simultaneous task alone. Furthermore, those *S*s of the latter group who reached the criterion on the simultaneous problem were a selected group of *S*s in that they were the four who had managed to learn under the successive condition when it was presented first.

Discussion

At first glance, these data lend support to the contention that it is more difficult to learn a successive discrimination than a simultaneous discrimination when the discrimination is difficult. However, a comparison of the performance of the group that learned to respond or inhibit a response (group III) with that of group I, which learned to press buttons to the left or right in the presence of positive and negative stimuli, indicates clear superiority for the former and emphasizes the importance of response variables in such learning. Grice, in a study with rats (2), found no difference in performance between the group which received the stimuli simultaneously and the group which received the stimuli successively. It is important to note that in Grice's successive presentation condition only a single stimulus was presented at a time and *S* was reinforced for running to the positive stimulus and not reinforced for running to the negative stimulus. This condition was somewhat similar to the approach task for group III of this experiment. The fact that groups II and III differed more in their performance on the first task in this experiment than did Grice's two groups may possibly be explained by his use of the more sensitive latency measure of discrimination. His criterion for a correct response was a latency to the positive stimulus less than, or a latency to the negative stimulus greater than, the median latency for a 10-trial block. The criterion in the present experiment was pressing or not pressing the button. Thus, it might be expected that if a latency criterion had been used in this experiment there would have been an even smaller difference between groups II and III. In the final analysis, then, this study tends to support the hypothesis of no difference between the two presentation conditions even with a difficult discrimination, if response variables are relatively well controlled, and the stimuli vary along a simple dimension.

The transfer data are somewhat in conflict with Grice's data which showed greater transfer from successive discrimination to simultaneous discrimination than for transfer in the other direction. The differences between groups are relatively slight, however, in both the Grice experiment and this one.

Unfortunately the subject pool was not large enough to include a fourth group that would have received two large or two small blocks on each trial, with the buttons continued in their left-right orientation. This condition would have maintained the block-button spatial relationships constant for both the simultaneous and successive conditions. It also would have approximated the condition used by MacCaslin (6). The slight difference in instructions between the simultaneous and successive conditions would have remained. Also, the latter successive presentation condition still calls for two responses, e.g., large-right, small-left, as compared with the large-approach of the simultaneous presentation. Thus, it is expected that the performance of such a group would not differ in any major way from group I of this experiment.

The fact that so many *S*s in the present study failed to stay with the successive situation when it came first indicates that the discrimination was difficult. These data may also be taken as another indication (3, 4) that one should be careful about accepting failure to learn as demonstrating lack of readiness to learn, when it may only mean less than optimal training conditions.

SUMMARY

This experiment was devised to evaluate the effect of (a) the simultaneous and successive methods of presenting stimuli and (b) the type of response required in the successive discrimination on the acquisition of a difficult discrimination. Also, transfer among methods was assessed. Three groups of five-year-old children served as *S*s. The simultaneous discrimination, involving pressing the button closer to the positive stimulus which appeared on the left or right, was easier than the successive discrimination when *S*s were required to push a button on the left to the positive stimulus and on the right to the negative. There was not a significant difference, however, between the simultaneous and successive discriminations when *S*s in the latter situation were required to push a single button when the positive stimulus was present and not to respond when the negative stimulus was present. Transfer to the two-response successive discrimination was best after learning under the simultaneous condition. Only four of the nine *S*s learned the discrimination when they were trained first under the successive condition with the left-right response.

REFERENCES

1. BITTERMAN, M. E., & WODINSKY, J. Simultaneous and successive discrimination. *Psychol. Rev.*, 1953, 60, 371-376.
2. GRICE, G. R. Visual discrimination learning with simultaneous and successive presentation of stimuli. *J. comp. physiol. Psychol.*, 1949, 42, 365-373.
3. JEFFREY, W. E. Variables in early discrimination learning: I. Motor responses in the training of a left-right discrimination. *Child Develpm.*, 1958, 29, 269-275.
4. JEFFREY, W. E. Variables in early discrimination learning: II. Mode of response and stimulus difference in the discrimination of tonal frequencies. *Child Develpm.*, 1958, 29, 531-538.
5. LOESS, H. B., & DUNCAN, C. P. Human discrimination learning with simultaneous and successive presentation of stimuli. *J. exp. Psychol.*, 1952, 44, 215-221.
6. MACCASLIN, E. F. Successive and simultaneous discrimination as a function of stimulus-similarity. *Amer. J. Psychol.*, 1954, 67, 308-314.

TRANSPOSITION IN THE INTERMEDIATE-SIZE PROBLEM BY PRESCHOOL CHILDREN[1]

HAYNE W. REESE*
University of Buffalo

Several investigators (e.g., 2, 3) have reported that children tend to transpose in the intermediate-size problem if they possess a concept, such as "middle-sized," that can serve as a mediational mechanism, but otherwise tend not to transpose. However, there is some evidence (1, 4) that children who do not possess such a concept also tend to transpose. The present experiment is a further investigation of this problem.

METHOD

Subjects

The *S*s were 36 children from the Iowa Child Welfare Research Station Preschool Laboratory and the Parents Preschool of Iowa City, with CAs from 3 years, 7 months, to 5 years, 4 months. Six additional *S*s did not complete the experiment because of failure to reach the learning criterion.

Apparatus

The apparatus consisted of a tray and a screen which could be rotated to hide or reveal the surface of the tray. The stimuli were two sets of wooden blocks, one set used in a pretest and the other used in training and transposition tests. The pretest stimuli were round and painted light green, and the other stimuli were square and black. Within each set each stimulus bore a 1.6 to 1 area-ratio to the next smaller stimulus. The areas varied from about 9 sq. in. to about 96 sq. in. A hole in the bottom of each block could be used to hide the incentive, a marble.

Procedure

At least 10 days before training began, a pretest was given to determine whether *S* possessed the concept of intermediate size. Three round, green stimuli were presented on each of six trials, using different absolute sizes of stimuli on different trials but retaining a constant area-ratio. The *E* named the sizes of the large and small stimuli, and *S* was required to emit a word designating the size of the middle-sized stimulus. The *S*s who emitted an appropriate word on at least five of the six trials were assigned to a "concept" group, group C. The other *S*s were given six further trials in which they were asked to point to the middle-sized stimulus. The *S*s who pointed incorrectly four or more times were assigned to a "no-concept"

---

* Department of Psychology, University of Buffalo, Buffalo 14, N.Y.

[1] The experiment was conducted while the writer was a research assistant at the Iowa Child Welfare Research Station. The writer is indebted to Mrs. Joan Danskin Kemble, who served as *E* for the pre- and posttests, and to Dr. C. C. Spiker for his many helpful suggestions.

Child Develpm., 1961, 32, 311-314.

group, group NC. The remaining *Ss* were excluded from the experiment since their classification was doubtful. Five *Ss* were excluded. There were 18 *Ss* in each group.

The three largest square, black stimuli were used during training trials, and responses to the middle-sized stimulus were rewarded. Training was continued to a criterion of eight correct responses in 10 consecutive trials, using a modified correction procedure in which *E* showed *S* which stimulus he should have chosen, without giving *S* the marble. Six *Ss* were excluded for failing to reach criterion within 54 trials.

Test trials were given immediately after *S* reached the training criterion. Three six-trial tests were given, using the square, black stimuli. In the Near test the stimuli were one step removed from the training stimuli; in the Middle test, two steps removed; and in the Far test, three steps removed. Six subgroups were formed, each taking the three tests in a different one of the six possible sequences. All responses were rewarded during the test trials.

At least one day after the experimental session *Ss* were given a posttest, identical to the pretest, to obtain further information about *Ss*' possession of the concept of intermediate size.

RESULTS

Training Trials

The mean number of trials through criterion was 17.1 (SD = 11.1) for group C, and 24.7 (SD = 15.5) for group NC. The difference was not statistically significant ($F = 2.39$; $df = 1, 24$; $p > .10$).

Test Trials

Table 1 presents the mean numbers of responses to the three stimuli in each of the three tests. The frequency of responses to the middle-sized stimulus decreased with increasing distance between the training and test stimuli and decreased more in group NC than in group C. The Distance effect was reliable ($F = 15.39$; $df = 2, 48$; $p < .001$), but the effect of the Concept variable was not significant ($F < 1.00$), and the interaction between the Distance and Concept variables was not significant ($F = 1.52$; $df = 2, 48$; $p > .20$). There was significantly more transposition in the Near test than in the Middle test ($t = 4.46$)[2] and in the Far test ($t = 5.09$), but no significant difference between the latter two tests ($t < 1.00$).

The frequency of responses to the largest stimulus increased with increasing distance between the training and test stimuli and increased more in group NC than in group C. The Distance effect was reliable ($F = 15.19$; $df = 2, 48$; $p < .001$), but the Concept effect and the Distance by Concept interaction were not significant (each $F < 1.00$). There were significantly fewer responses to the largest stimulus in the Near test than

---

[2] The error terms of the *t* tests were based on the within-*Ss* error terms of corresponding analyses of variance, and the *df* of *t* are the same as the *df* of the within-*Ss* error term.

TABLE I

MEAN NUMBER OF RESPONSES TO THE THREE STIMULI ON TEST TRIALS

Stimulus	*Group*	TEST *Near*	 *Middle*	 *Far*
Small	C	1.1	1.0	1.3
	NC	0.9	1.5	1.8
Middle-Sized	C	2.9	1.6	1.3
	NC	3.4	0.7	0.5
Large	C	2.0	3.4	3.4
	NC	1.7	3.8	3.6

in the Middle test ($t = 4.89$) and in the Far test ($t = 4.66$), but there was no significant difference between the latter two tests ($t < 1.00$).

The frequency of responses to the smallest stimulus increased with increasing distance and increased more in group NC than in group C, but none of these effects was reliable (Distance effect: $F = 1.66$; $df = 2, 48$; $p > .20$. Concept effect: $F < 1.00$. Distance × Concept interaction: $F < 1.00$). However, the Distance variable interacted significantly with Subgroups ($F = 3.02$; $df = 10, 48$; $p < .005$). The frequency of responses to the smallest stimulus increased with increasing distance except in the two subgroups given the middle test first; in the subgroup given the sequence Middle-Far-Near the trend was reversed; and in the subgroup given the sequence Middle-Near-Far there was only one response to the smallest stimulus out of 54 responses. These simple effects were not analyzed statistically.

Posttest

Two *S*s, one from group C and one from group NC, missed the posttest. All remaining group C *S*s were also classified as group C on the basis of posttest scores. Of the 17 group NC *S*s who took the posttest, 12 were also classified as group NC on the basis of posttest scores; four were classified as group C; and one was classified as "doubtful." The pre- and posttest classifications were significantly related ($\chi^2 = 16.93$; $df = 1$; $p < .001$). The *S*s were reclassified on the basis of their posttest scores, but the group differences were smaller than those in Table 1 and were not analyzed statistically.

DISCUSSION

The *S*s tended to transpose on the Near test, whether or not they possessed the concept of intermediate size, more than on the other two tests. This finding supports the theory (4) that the probability of transposition is an inverse function of the discriminability of the training and test sets of stimuli, on the assumption that the discriminability increases as the number of steps between the training and test sets of stimuli increases.

The frequency of responses to the largest stimulus, which was the most similar to the positive stimulus of the training trials, increased as the frequency of transposition decreased, suggesting that, when *S*s discriminate between the training and test sets of stimuli, they respond on the basis of absolute size.

SUMMARY

Preschool children were divided into a Concept group and a No-Concept group on the basis of a pretest in which they were required to name the middle-sized of three stimuli differing in area. At least 10 days after the pretest, they learned to choose the middle-sized of three other stimuli and were then given test trials in which the stimuli were one step (Near test), two steps (Middle test), and three steps (Far test) removed from the training stimuli.

There was no significant difference between the Concept and No-Concept groups in trials to criterion in the training phase and no significant difference in performance on test trials. Distance effects were obtained, with significantly more transposition and fewer responses to the largest stimulus on the Near test than on the Middle and Far tests, which did not differ significantly from each other. The distance effects did not interact significantly with the concept variable.

The data support the theory that the probability of transposition is an inverse function of the discriminability of the training and test sets of stimuli and suggest that, when *S*s make the discrimination, they respond on the basis of absolute size.

REFERENCES

1. REESE, H. W. Transposition in the intermediate size problem as a function of area-ratio of stimuli. Unpublished master's thesis, Univer. of Texas, 1955.
2. SPIKER, C. C., GERJUOY, I. R. & SHEPARD, W. O. Children's concept of middlesizedness and performance on the intermediate size problem. *J. comp. physiol. Psychol.*, 1956, 49, 416-419.
3. SPIKER, C. C., & TERRELL, G., JR. Factors associated with transposition behavior of preschool children. *J. genet. Psychol.*, 1955, 86, 143-158.
4. STEVENSON, H. W., & BITTERMAN, M. E. The distance-effect in the transposition of intermediate size by children. *Amer. J. Psychol.*, 1955, 68, 274-279.

A COMPARISON OF SOCIAL BEHAVIOR IN NORMAL AND HYPERAGGRESSIVE PREADOLESCENT BOYS[1]

ALLEN T. DITTMANN* *and* D. WELLS GOODRICH
National Institute of Mental Health

This paper presents data on the free social behavior of two groups of preadolescent boys of about the same age. The first is a normal group, selected so as not to include any gross psychopathology, and the second is an in-patient group selected as ego-disturbed, hyperaggressive boys as described by Redl and Wineman (7). Both were studied under the same conditions of living arrangements, program, and adult staff. The patients were part of the same group on which observations have been reported elsewhere (5, 6), but the data to be reported here were collected about a year earlier, when the first boys of the group had just been admitted. The normal group was not a control group in the sense of being precisely matched on a number of variables, but were boys admitted partly for staff training and partly for research purposes before the patient group arrived.

The first purpose of the research is to confront theoretical formulations and case reports of hyperaggressive children with actual behavior of these children. Such patients have been extensively described by Redl and Wineman (7) as children whose ego organizations lack control and whose chief reactions take the form of outbursts of rage. The examples given in the books certainly support the formulations of the pathology, and the case materials of our present patients are also replete with illustrations of unmanageable rage in school, family, community. Such case descriptions, however, are necessarily overinfluenced by the more outstanding and colorful

---

* Laboratory of Psychology, National Institute of Mental Health, National Institutes of Health, U. S. Public Health Service, Department of Health, Education, and Welfare, Bethesda 14, Maryland.

[1] This study was carried out in the Child Research Branch of the National Institute of Mental Health, and the authors wish to thank Dr. Fritz Redl, then Chief of the Branch, for his support in collecting the observations on which the study is based.

Child Develpm., 1961, 32, 315-327.

incidents in which the child has become involved. While it is clear that he has been aggressive enough to have case records written about him in the first place, still it is not clear how much of his aggression is pathological. By systematic sampling of a wider range of behavior, in addition, the data will give a more complete picture of areas of ego strengths and defects. Research which attempts some representation of the child's daily life activities has long been needed as a check on clinical case material.

The second and more obvious purpose of the present study is that of obtaining baseline data on the behavior of normal preadolescent boys against which to cast the observations of our patients. With information on normals, we can draw conclusions, not only about how much of our patients' behavior is pathological, but how much of it represents the islands of nonpathology which treatment agents try to exploit. Data on normals can also help the staff to know which of the new behaviors which appear in the course of treatment can be interpreted as signs of progress. Studies of normal children reported in the literature are of little help for our purposes, since the settings in which the children have been observed, i.e., camps, boys' clubs, and the like, have been so different from the in-patient treatment situation in which our patients find themselves.

Since the present work was an early study in the program as a whole, there were a number of areas in which controls could not adequately be worked out. First, the number of children was greater in the normal group than in the patient group. Second, the normal group was somewhat older and more homogeneous in age than the patient group. Expectations of how these areas might affect the ego controls of the children are opposite: the younger patient group would be expected to have less control than the older normals on the basis of age alone, while the smaller group of patients (with the same size of child care staff) would be expected to provide fewer opportunities for outbursts of rage.

METHOD

The groups under study lived on the nursing unit of the Child Research Branch of the National Institute of Mental Health. This unit is on the fourth floor of the Clinical Center of the National Institutes of Health, arranged so that four bedrooms are at one end and the dayroom for indoor activities and dining area are at the other. Since the observations in this study were made during summer and early fall, much activity took place out of doors, in a private enclosure just outside the building, and on trips away from the grounds. The four social settings in which the children were observed are described below: (a) Meals, mostly in the nursing unit, but occasionally outside; (b) Group Activities, in which the boys as a group, or a majority of the boys, were together, either engaged in a single game or divided into subgroups which were fairly close to each other; (c) Arts and Crafts, where all children were making the same thing, such as model planes, Indian headpieces, and the like, with child care staff or occupational therapists available for help and instruction; (d) Snacks, the time after showers, and just before bedtime. About two-thirds of the observations were of the Group Activities setting, which contained a great deal of social interaction.

Observations were done by the method of Rauch and Rauch [described in detail by M. D. Rauch (4)]. Briefly the observer was nonparticipant in the situation and observed the total group of children, plus the child care staff on duty at the time, for a period of time determined by his own judgment of how much he could remember. The average time was a little less than 10 minutes and varied according to the amount of activity and the cohesiveness of sequences of interaction: the time was apt to be short where the children were very active or where there was a series of unrelated interactions which the observer felt would be confusing to remember in sequence. Following the observation, the observer went to an office where an interviewer would try to elicit as complete a picture as possible of the interpersonal events which transpired during the period of observation. This interview was recorded, and the recording served as the basic protocol of coding.[2]

Each observation was coded for interactions on the Freedman-Leary-Ossorio-Coffey Circle of Ego-oriented Interpersonal Mechanisms (3), using the 16 categories which that system provides and the three degrees of "intensity." The categories are various combinations of the two dimensions of affect (friendly to hostile) and status (dominant to submissive), while the intensity parameter includes uninvolved, appropriately involved, and inappropriately intense levels of behavior. The reliability of the coding in this setting, and of observations as seen through the codes, is reported in detail elsewhere (2). Briefly, the coding of individual interactions is low in reliability, though clearly beyond chance expectations. The reliability of profiles made up of these codings, however, is very high. The ratings of intensity are of relatively low reliability, suitable only for studying group differences. Coding was done only on the Public or Interpersonal Level of the system; our results do not include any data on the Private or Symbolic levels. Profiles were made up for each child, keeping his interactions with other children separate from his interactions with adults of the child care staff and keeping the degrees of intensity also separate. Profiles were also made up of the adults' interaction with the children since we were interested in how constant the adults were in dealing with the two groups of children.

---

[2] The method reported here differs from the work of Raush *et al.* (5, 6) in three respects, and the findings are therefore of questionable comparability: (a) the groups were observed as a whole in this study, whereas in later work the observer concentrated in each observation on the behavior of one child plus other people he interacted with—a more feasible task for an observer; (b) rather than the four social settings reported on here, six were used by Raush *et al.* (5, 6) in a balanced setting-by-child design; (c) the interview of the observer, while useful as a training device, was found not to elicit more interactions per observation (2) and was therefore not continued in later studies.

Results reported here are of total profiles of all children in each group, broken down by degrees of intensity of interaction. There were in all 49 observations of the normal group, comprising 1674 interactions, and 26 observations, 764 interactions, of the patient group.

SUBJECTS

The normal group consisted of eight boys, four white and four Negro, average age 10 years, 3 months; range 9 years, 8 months to 11 years, 2 months. All had just completed fifth grade in public schools. The initial selection was made by school personnel, who based their judgment chiefly on the children's popularity with other children, being careful not to select outstanding leaders or followers, or very bright children, or children especially eager to conform to adult demands for politeness and studiousness, or children either with many behavior problems or with absolutely none. The children did not know each other before their stay at the Clinical Center. After this initial selection was made, our staff observed the children on the playground and in the classroom at school, and, in the case of those who were accepted on this basis, interviews were held with the parents. The orientation for them was that we were interested in having a group of normal children to study in a situation which would later be used to study disturbed children, that we would run a program similar to summer camp activities. The parents were also told that we would like to call on them in the future for brief visits to the Clinical Center. Despite some understandable initial uneasiness on the part of some of the parents, the response was very favorable. The children were formally admitted as normal control subjects and stayed four weeks. Observations were carried out during the second and fourth weeks. The first week was left free to allow the children to get used to the Clinical Center life and to become acquainted with the observers' presence as nonparticipants. These children differentiated the observers very readily from the child care staff and after a few days did not ask observers for help in tasks or for participation in games, even though they were always very friendly to the observers.

The disturbed group consisted of four boys, average age 9 years, 5 months; range 8 years, 2 months to 10 years, 2 months. These were the first four of six boys who have made up the subjects for the main long-term treatment study of the Child Research Branch. Since their selection was based on more stringent criteria, they could not all be admitted at the same time. They are described in some detail in Redl and Wineman (7) and Raush *et al.* (5).

In the interests of comparability of data we tried to collect observations on the patients in the second and fourth week after their admission. The observations extended over a somewhat longer period, however, partly because the children's stay was not brief and partly because extra care had to be exercised to insure that the observers did not interfere with the ongoing activity program for the boys. They were highly distractable, and the addition of another adult, even a nonparticipating one, could mean a new outburst of rage. It took many weeks for the boys to recognize the observer role as such. During the period covered by this report, the program for this patient group was almost identical with that arranged for the normal group. Neither school nor regular psychotherapeutic interviews had yet been instituted at that time.

RESULTS

Method of Presenting Results

Those findings which are best presented as profiles of the groups' interactions are broken down by degree of intensity and by object of interaction. Intensities are separated for two reasons: (a) we wished to compare the groups in nonextreme behavior in order to be able to assess any qualitative differences in preferred interpersonal response; (b) since there is considerable difference between the groups in intensity of interactions, we wished to control intensity by examining the profiles within each level of intensity. Profiles are separated into interactions toward peers and interactions toward adults. The profiles themselves have been drawn on polar coordinate graphs in which the radial distance represents percentages of frequencies such that the total percentage for each profile equals 100. The radii of the graphs are arranged so that areas are proportional: the area in a sector representing 20 per cent, for example, includes twice that of a sector representing 10 per cent.

Degrees of Intensity of Interactions

The distribution of interactions by degrees of intensity for the two groups is presented in Table 1. The χ^2 test of these frequencies is 135.16, with 2 degrees of freedom, according to which the hypothesis of randomness may be rejected with greater than .001 confidence.

TABLE 1

INTENSITY OF INTERACTIONS IN NORMAL AND HYPERAGGRESSIVE GROUPS

	NORMAL GROUP		HYPERAGGRESSIVE GROUP	
Degree of Intensity	*Frequency*	%	*Frequency*	%
1. Noninvolved behavior, appropriate to situation	267	16.0	86	11.3
2. Involved behavior, appropriate to situation . . .	1304	78.0	484	63.3
3. Involved behavior, inappropriately intense . . .	103	6.0	194	25.4
	1674	100.0	764	100.0

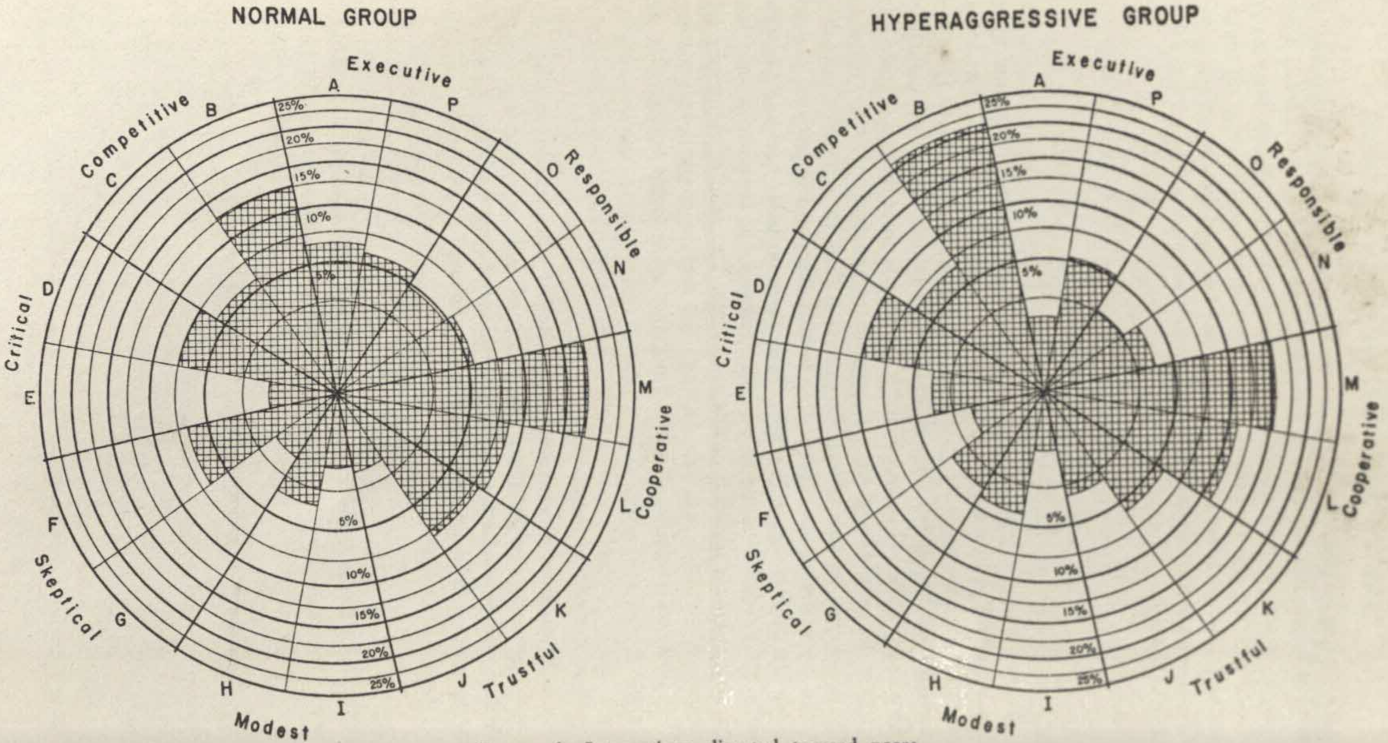

Figure 1—Interactions directed toward peers.

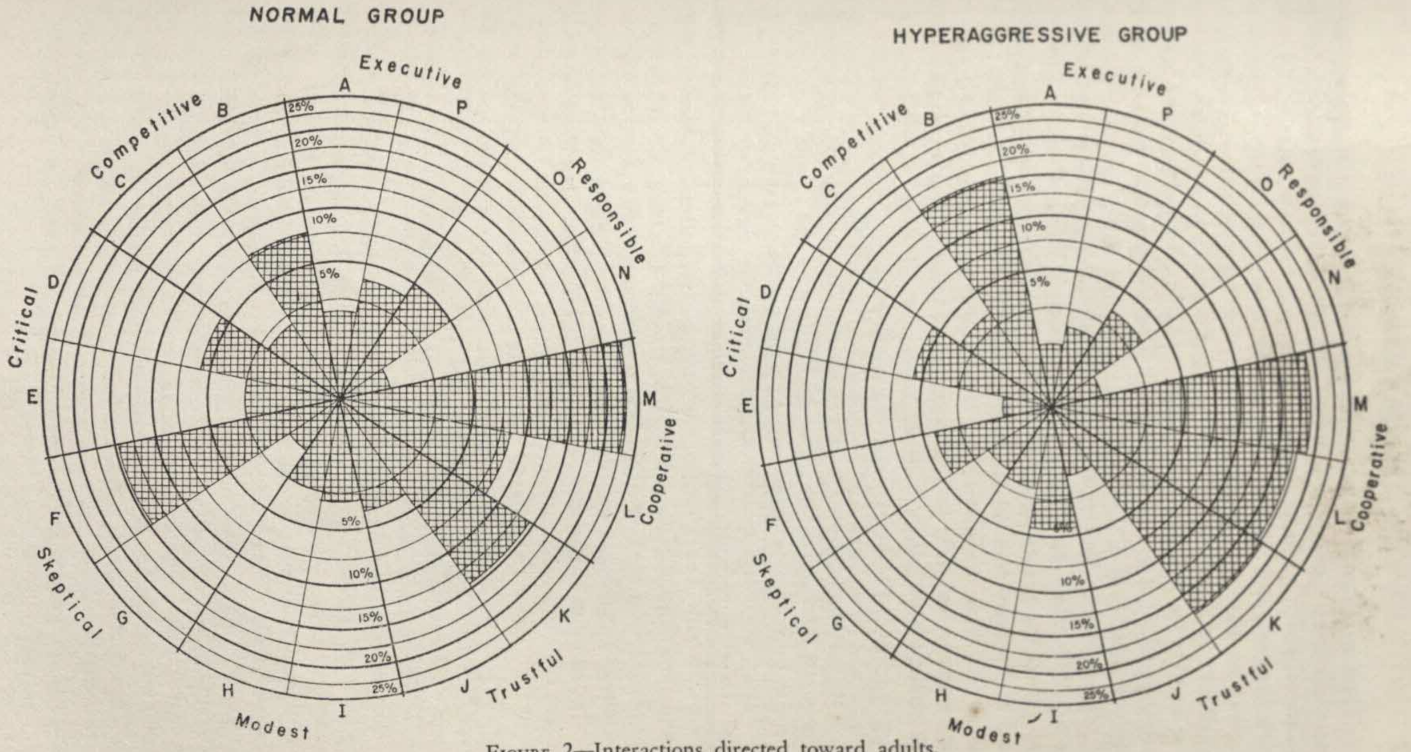

FIGURE 2—Interactions directed toward adults.

There is certainly nothing surprising in this finding. The greatest difference between the groups as seen in Table 1 is in the inappropriately intense behavior, which made up a quarter of the patient group's interaction in the time observed, and this in a situation where there were plenty of adults, all bent on establishing a psychologically hygienic atmosphere and on building an activities program in which inner controls could grow. Clearly this finding answers the first question of this research: referring agencies have not overemphasized a few isolated instances of high intensity behavior in describing these boys.

The difference between the two groups on this variable is also reflected in the differential reaction to them by the child care staff. Analysis of the intensity of their reactions toward the children also yields differences significant at the .001 level of confidence, though the differences are not so striking as those in Table 1 since the frequency of inappropriately intense interactions is small for both groups: 3 per cent to the normals and 9 per cent to the patient group.

Comparison of Interaction Profiles of the Two Groups

The profiles of the two groups at intensity 2, the involved, situationally appropriate interactions, are presented in Figure 1 and 2, which are profiles of interactions of the children toward peers and toward adults, respectively.

Taking responses to other children first, the similarities between the profiles are immediately apparent, and yet their distributions are significantly different. The χ^2 test of the frequencies of these profiles is 37.07, significant at almost the .001 level. The greatest difference lies in the upper half of the circle: the normal children are friendly in their dominant behavior toward their peers, and the hyperaggressive children most hostile, as is expected from the selection of the groups. A difference which is not so obvious is in the variability of responsiveness. Whereas the patient group's use of the various categories of the Circle are either extremely high or extremely low, the normals' distribution is more evenly spread out on the graph, less characterized by peaks and dips. Thus we see that there is greater differentiation of interaction among the normals and a tendency toward stereotypy among the patients.

The patients show a considerably greater tendency to express themselves in a boastful, challenging, or provocative manner. The normals demonstrate a definitely greater capacity for all types of interaction, for a wider variety of interpersonal expression, and especially for giving appropriate orders and acting helpfully. These leadership behaviors undoubtedly account for the fact that the normal boys required much less adult help in initiating and maintaining constructive group activities and were much more self-directed as a group. Here also is expressed in quantitative terms what we know clinically about the ego of the severely disturbed child, namely that there is less differentiation of capacities for dealing with relationships.

Looking at Figure 2, we see greater apparent differences between the profiles of the groups' interactions with adults than we saw among peers.

The χ^2 test yields a figure of 61.18, which exceeds the .001 significance level. There are differences in all quadrants, with the normals showing more interactions in the friendly-dominant and hostile-submissive quadrants than the patients. The single category which contributes fully a third of the χ^2 total is F (complains, rebels), in which the normal children contribute far more than expected by chance. This finding is not the result of overweighing one or two childrens' profiles: F was in the first or second rank for five out of the eight children, the lowest rank being 3.5. It should be noted that, even if F had contributed nothing to the total χ^2, the difference between the two profiles would still be significant at the .001 level. In the matter of variability of responsiveness in interactions toward adults, the hyperaggressive children are again more stereotyped than the normals, though the difference is not as apparent as in the case of interaction toward peers.

Here again the data do not surprise us, in the light of clinical studies. The inability of our patients to express hostile wishes towards adults in a passive mode—as would be appropriate in our culture for the latency child—is based upon their need to maintain active control over any hostile or anxiety-laden aspect of a relationship. Since relationships for these children are usually experienced with the anticipation of rejection, punishment or damaged self-esteem, it becomes especially dangerous to remain passive during a moment of disturbance with another person, especially an adult (1).

Inappropriately Intense Interactions

These are presented in Tables 2 and 3. These are tabulated by octants of the Circle because of the low frequencies in several of the 16 individual categories. In computing χ^2 tests for these tables, even further collapsing is necessary to bring expected frequencies up to workable numbers.

TABLE 2

INAPPROPRIATELY INTENSE INTERACTIONS OF NORMAL AND HYPERAGGRESSIVE GROUPS TOWARD PEERS

	NORMAL GROUP		HYPERAGGRESSIVE GROUP	
Interaction Categories	*Frequency*	*Per Cent*	*Frequency*	*Per Cent*
Bossy (A-P)	8	12	5	6
Swaggering (B-C)	19	29	17	19
Combative (D-E)	16	24	28	31
Nagging (F-G)	8	12	9	10
Cringing (H-I)	3	5	5	6
Clinging (J-K)	1	2	5	6
Effusive (L-M)	9	14	17	19
Indulgent (N-O)	2	3	3	3
Total	66	101	89	100

TABLE 3

INAPPROPRIATELY INTENSE INTERACTIONS OF NORMAL AND HYPERAGGRESSIVE GROUPS TOWARD ADULTS

	Normal Group		Hyperaggressive Group	
Interaction Categories	*Frequency*	*Per Cent*	*Frequency*	*Per Cent*
Bossy (A-P)	6	16	1	1
Swaggering (B-C)	3	8	29	28
Combative (D-E)	16	43	36	34
Nagging (F-G)	3	8	11	10
Cringing (H-I)	0	0	4	4
Clinging (J-K)	5	14	7	7
Effusive (L-M)	3	8	16	15
Indulgent (N-O)	1	2	1	1
Total	37	99	105	100

The difference between the profiles in Table 2 can be expected about 30 per cent of the time by chance, but those in Table 3 less than 0.1 per cent. The main differences in the behavior toward adults lie in the proportions of responses in the friendly-dominant quadrant (A-P and N-O) where the normal group shows high frequencies and to some extent also in the hostile-passive quadrant (F-G and H-I) where the hyperaggressive group contributes more than its expected share. The profile differences between intensity 2 and 3 behavior are statistically significant for both groups although the differences are more marked for normals and patients alike in the case of their interactions with adults.

Of particular interest theoretically is this finding that these patients tend to express their most intense psychopathology more with adults than with their peers. Redl and Wineman (7) and others have long recommended that treatment with these children should begin within a peer group activity setting since these children find it easier to form relationships with peers. With adults, particularly with interested warm adults, intense anxieties about the fantasied dangers believed by the child to be inherent in such trusting dependent closeness are mobilized and acted out. In part, the difference in locus for the expression of psychopathology in these patients can be accounted for by their failure in superego maturation. Normally in this developmental phase, the child identifies sufficiently with the "good authority" aspects of the parent to displace more and more of his anxieties into the peer-group and sibling relationship arenas. (This is borne out in the normals' greater amount of inappropriately intense behavior to peers than adults.) With hyperaggressive, or other children with severe developmental personality defects, the tendency is to maintain the experience of anxieties directly with the parental surrogate.

Uninvolved Low Intensity Interactions

Behavior at level 1 of the intensity scale is rare in the patient group, and comparisons are therefore impossible to carry out in parallel to the findings reported above. The profiles of interaction of the three intensities are, however, very similar in both groups, with the interesting exception among the normals of levels 1 and 2. The hostile-dominant behavior of the normals tends to be uninvolved, while their passive-friendly behavior tends to be involved.

Adult Interaction toward the Two Groups

The profiles of the child care staff's interactions to the two groups of children have been computed only at intensity 2, the frequency of their intensity 3 behavior being too small to handle statistically. In general, about two-thirds of their behavior falls into the friendly-dominant quadrant of the circle, as their role dictates. Indeed, about a fifth of their interactions fall into each of the categories A (commands, directs) and O (offers, helps). When the adults' behavior toward the two groups of children is compared, the difference is significant at about the .05 level. The chief differences lie in the most friendly and most hostile octants of the circle, and in both of these their behavior toward the normals exceeds chance expectations. This finding is in line with their spontaneous comments on the differences between their tasks in dealing with the two groups: with the normals the adults could relax, be more spontaneous in reacting to the children—these healthy children did not keep grudges and did not hold past favors over one's head to try to extract future privileges. With the patient group, on the other hand, the child care staff had to be cautious, because the children were so changeable and panic or tantrums were frequent enough to lead the adults to exercise tighter control, not only on the children, but also on their own expressiveness.

DISCUSSION

In general, both quantitative and qualitative differences between our groups of normal and hyperaggressive preadolescents confirm past clinical observations and theoretical constructions—the hyperaggressive boys are indeed more aggressive, more stereotyped in their behavior, and less controlled. We should note, however, in looking at the profiles, that there is also considerable similarity in their interaction profiles. We would hope that our research would show the patients' resemblance to other boys, despite their differences from normal ones their age.

There are a few surprises in the data. For example, clinicians familiar only with the literature on these disorders, literature which expresses the rage and destructive content of these children's behavior, may be surprised at the large amount of often intense warmth and friendliness (coded M) in the patient group profile. True, the child care workers who lived with them found this friendliness difficult to trust and did not respond to it as freely as they did with the normal group. But it is friendliness just the same and represents one of the areas of possible contact with these children, like the children's need for activity and vigilance about fair play, which treatment programs seek to exploit. But the patients' relative lack of helpful and supporting warmth (coded O and N), as well as the decided absence of true leadership qualities (coded A), as compared with the normals, must be considered an ego defect in its own right, not just an "acting out of unconscious conflict." Even without the aggressive components of their behavior, the lack of these qualities would make it hard to live with these children.

Perhaps the most unexpected finding is the relatively low frequency of overt attacking behavior (coded E) in these patients—very similar to the amount shown by the normals. There were, to be sure, marked differences in the intensity of attack in the two groups when they did display direct aggression, but pure attack was rare. The fact that many of the observations were taken during group activities, a setting which seemed often to evoke competitiveness, narcissistic wounding, and rage in our patients, makes the low frequency of direct attack doubly surprising. Perhaps the best explanation is the degree to which the adult child care workers were conscious of their therapeutic responsibility to prevent and absorb the hatred of these boys. Considerable time was devoted to designing a program which could drain off aggression in structured ways, through games and other activities.

SUMMARY

Two groups of preadolescent boys, eight well-functioning normals and four hyperaggressive patients, were studied by observation of free social behavior in the identical living situation, adult personnel, and programmed activities. While there is similarity between the groups, the differences in a number of areas were striking:

1. The hyperaggressive group showed more intensity of interaction than the normals.

2. In quality of normally involved interaction toward peers, the patients were more hostilely dominant and less dominant in a friendly, supporting way than the normals.

3. Toward adults the patients were also more hostilely dominant and less friendly dominant than the normals. The normals, on the other hand, were more hostilely submissive to the adults than were the patients.

4. In inappropriately intense behavior toward peers, the two groups were very similar. Toward adults, however, there were strong differences: intense reactions of the normals were at once more friendly-dominant and more hostilely-submissive than the patients.

5. The patients' behavior was rather stereotyped, with a few categories accounting for a large proportion of their interactions, while the normals used a wider repertoire of behavior.

6. Adults' behavior toward the two groups was much less different than the groups were from each other, most of their behavior being role-appropriate helping, supporting, and leading. The small amount of inappropriately intense interactions they displayed was directed more toward the patients than toward the normals. As regards appropriately involved behavior, the adults were more frankly hostile and friendly toward the normals and more controlled in dealing with the patients.

REFERENCES

1. BLOCH, D. A. The delinquent integration. *Psychiatry,* 1952, 15, 297-303.
2. DITTMANN, A. T. Problems of reliability in observing and coding social interactions. *J. consult. Psychol.,* 1958, 22, 430.
3. FREEDMAN, M., LEARY, T., OSSORIO, A., & COFFEY, H. The interpersonal dimension of personality. *J. Pers.,* 1951, 20, 143-161.
4. RAUCH, M. D. An investigation of relationships between social behavior and psychiatric status in hospitalized mental patients. Unpublished doctoral dissertation, Univer. of California, 1953.
5. RAUSH, H. L., DITTMANN, A. T., & TAYLOR, T. J. The interpersonal behavior of children in residential treatment. *J. abnorm. soc. Psychol.,* 1959, 58, 9-26.
6. RAUSH, H. L., DITTMANN, A. T., & TAYLOR, T. J. Person, setting, and change in social interaction. *Hum. Relat.,* 1959, 12, 361-378.
7. REDL, F., & WINEMAN, D. *Children who hate.* Free Press, 1951.

INTERACTION BETWEEN VERBAL AND NONVERBAL BEHAVIOR[1]

O. IVAR LOVAAS*[2]
University of Washington

Experimental investigations on the relation between verbal and nonverbal behavior date back as far as Lehmann's study (5) on the effect of verbal labels upon discrimination. Since then, research has been oriented around discrimination learning, generalization, transfer, transposition, problem solving, etc. Thus, previous research on the relation between verbal and nonverbal behavior has been frequently tied in with traditional areas of psychological inquiry where most often the nonverbal behavior has been employed as an index of the effect of verbal behavior on a hypothesized basic process, such as discrimination. The conceptual formulations designed to cover the empirical findings have been primarily concerned with the cue properties (or discriminative stimulus properties) of the two kinds of behavior—the two behaviors have been thought to interact in so far as one provides discriminative stimuli for the other. Esper states this interaction in a broad formulation: "the stimuli associated with each type of response are among the conditioned elicitors of the other . . ." (2, p. 446). A similar paradigm is employed by Kurtz and Hovland (4) to deal with the more specific problem of the effect of verbal behavior on discrimination; they apparently conceive of the effect as due to attending behavior elicited by the verbal response. Miller's (7) formulations on acquired equivalence and distinctiveness of cues similarly rely on the cue properties of the verbal behavior in effecting a change in nonverbal behavior. Indeed, all mediational hypotheses assume that it is the cue properties of the (hypothetical) verbal response that provide the "connecting link" to other behaviors.

In so far as one is concerned with the effect of verbal behavior on specific problems such as discrimination learning and concept formation, one can

---

* Gatzert Institute of Child Development, Department of Psychology, University of Washington, Seattle 5.

[1] This study was supported by grant M-2208 from the National Institute of Health, United States Public Health Service.

[2] The author expresses his gratitude to Professors Donald M. Baer and Sidney W. Bijou for their assistance, particularly on the conceptual aspects in this study. He is also grateful for the cooperation of the staff of the Nursery School, Gatzert Institute of Child Development.

Child Develpm., 1961, 32, 329-336.

perhaps adequately conceptualize the relations observed by considering only the stimulus properties of the verbal behavior. It is apparent that in so doing one limits oneself to only one of several possible interactions between the two behaviors. It is, for example, conceivable that the two classes of behavior could interact on the basis of having common reinforcing stimuli, common emotional states which influence either behavior, etc. An analysis in these terms becomes more appropriate when one deals with broader classes of verbal and nonverbal behavior, for example, if one deals with social behavior, as in the study to be reported.

The purpose of the present study was to determine the effect of strengthening one class of verbal responses on a class of nonverbal responses. A bar-pressing response reinforced by aggressive doll action was observed immediately after a verbal conditioning session during which one group of children was reinforced for emitting aggressive verbal responses and the other group was reinforced for emitting nonaggressive verbal responses.

Method

Apparatus and Subjects

The research was conducted in the laboratory of the Gatzert Institute of Child Development. One room, the observation room, was equipped with one-way mirrors and sound equipment. The *E* (experimenter) could present the various experimental treatments from this room without being observed by *S* (subject). The other room, the "playroom," contained the apparatus or toys with which *S* would play and a partition behind which *A* (an adult assistant) could sit.

There were three pieces of apparatus in the playroom. Two of these, the doll apparatus and the ball-toy, have been described earlier (6). Briefly, the doll apparatus, or "striking dolls," was arranged so that each depression of a lever in the box on which the dolls were placed would make one doll strike the other on the head with a stick. The ball-toy consisted of a ball within a cagelike structure; the ball could be flipped up and down inside the cage by depression of a lever at the base of the cage. Depression of these levers also activated electric counters. The two were placed on a table, enabling *S* to operate both simultaneously.

The third piece of apparatus was a "talk-box," a 15 by 12 by 12 inch wooden box, placed on a table at the opposite wall from the table containing the dolls and the ball-toy. A microphone, inside the talk-box, was wired into the observation room and enabled recordings of *S*'s verbalizations on a Gray Audograph recorder. Two dolls were fastened in a sitting position on top of the box. These dolls originally were similar in appearance, but one, "the bad doll," had been made very dirty; the other doll, "the good doll," was neat and clean. Reinforcement, in the form of small trinkets, could be delivered automatically to *S* through a chute emptying into a small tray on the side of the talk-box.

The *Ss* were children from the Institute's Nursery School. They were above average in intelligence (1). Nineteen children, age 3-5 to 4-7, were randomly selected from this group (but *Ss* known to be uncooperative were not asked). Five *Ss* were eliminated during the experiment, two because they expressed definite desire to leave, two others because they failed to respond to the verbal conditioning procedure (one could not be differentially conditioned, the other did not respond verbally at all), the fifth one because he had an identical ball-toy at home. The remaining 14 *Ss* had experiences as *Ss* in other experiments, but had no previous contact with the apparatus employed in this study. The *Ss* were seen in the morning while they were engaged in free play.

Procedure

A, a female adult who had visited the nursery prior to the experimental run and acquainted herself with the children, invited *S* to the playroom to "play some games." *S* was first made briefly acquainted with the playroom and then introduced to the striking dolls and the ball-toy. *A* pressed the lever on the dolls once and said, "When you press this bar (demonstrates) this (pointing) will happen. Now you do it (*S* is induced to press the lever five times). Now look here (pointing to ball-toy), when you press this lever (demonstrates) this (pointing) will happen. Now you do it (*S* is induced to press the lever five times). Now you play with the toys while I sit in my chair." *S* received a three-minute period of play, timed from the *S*'s initial depression of either lever. This period constituted the pre-experimental or operant level of behavior for the striking dolls and the ball-toy.

A returned on signal by *E* and placed the dolls and the ball-toy out of *S*'s view. She then seated him in front of the talk-box and said, "This is a talk-box; when you talk to this box, it will give you toys right here (points to reinforcement tray). Now see here are two dolls. This (pointing) is the good doll; this (pointing) is the bad doll. Say 'good doll' (if necessary coaches *S* to say, 'good doll'; this response is reinforced). See what you got; this is your toy; you can keep it. Now say, 'bad doll' (coaching if necessary; this response is also reinforced). See what the box gave you; this is your toy to keep. Now you sit here and tell the box all about the dolls; tell the box what is going to happen to the dolls." The dolls were included to give *S* discriminative stimuli for emitting aggressive or nonaggressive verbal responses. Additional instructions were needed for about one-fourth of the *Ss* equally distributed between the groups. If, after two minutes and again after four minutes, *S* had made no response that could be reinforced, *A* told him from behind her screen: "Say, 'good doll' (*S* is reinforced for saying this); now say, 'bad doll' (also reinforced); now talk about the dolls."

The *Ss* were divided into two groups of seven each. One group was reinforced for aggressive verbal behavior (AV-group), the other for verbal behavior other than aggressive (NAV-group). The criterion used for deciding whether a verbal response was aggressive or not was whether it was derogatory to the dolls or denoted a wish on the part of the child to damage or hurt the dolls. In fact, the only responses the AV-group emitted that fulfilled this criterion were: "bad doll," "dirty doll," and "doll should be spanked."

The *S*s from the AV-group and the NAV-group were matched in pairs on sex, age, number of reinforcements, and length of the verbal conditioning period. The magnitudes of the last two variables were determined by the AV-member of each pair who invariably was run first.

The verbal conditioning procedure necessitated initially reinforcing both aggressive and nonaggressive verbal responses (e.g., "good doll" and "bad doll") for both the NAV- and AV-groups to produce a high rate of verbal responding. A high rate of responding was defined as at least 12 reinforceable verbal responses within any two-minute period. A reinforceable verbal response was any word, phrase, or sentence. The criterion of high rate was reached from two to eight minutes after initiation of the verbal conditioning period. Once the high rate was reached, the AV-group became selectively reinforced for aggressive verbal responses while the NAV-group was reinforced for verbal responses other than aggressive ones. In the NAV-group, affectionate responses were reinforced as little as possible. If "good doll" reached a high rate, reinforcement was withheld in order to extinguish this response and produce other verbal responses. This condition was introduced to insure that any difference in subsequent play behavior between the two groups would not be a function of a contrast between "aggressive" and "friendly" verbal responding, but rather between aggressive and nonaggressive responding ("friendly" verbal responses could be incompatible with subsequent aggressive behavior, hence the data would be less clear than otherwise).

The length of the conditioning period was determined by *S* in the AV-group. The period was terminated after six minutes once a high rate of aggressive verbal responses was established (not less than 20 reinforceable aggressive verbal responses in a three-minute period). Thus, the conditioning period lasted anywhere from six to 14 minutes.

When the conditioning period was terminated, *A* told *S* that the "talk-box" did not work any more, helped him put his trinkets in a paper bag, and told him: "I will keep these toys for you while you play some more with this (pointing to the striking dolls) and this (pointing to the ball-toy) for just a few minutes more." The child was then allowed to play with the striking dolls and ball-toy for a period of four minutes. At the end of the four-minute period, *A* came out, gave *S* a thank you and his trinkets, and accompanied him back to his nursery school group.

A interacted with *S* in a matter-of-fact manner. She remained behind her partition except when she introduced apparatus. She did not interact with him unless he came over to her chair, at which point she said: "I'll sit here and you be over there for a few minutes more."

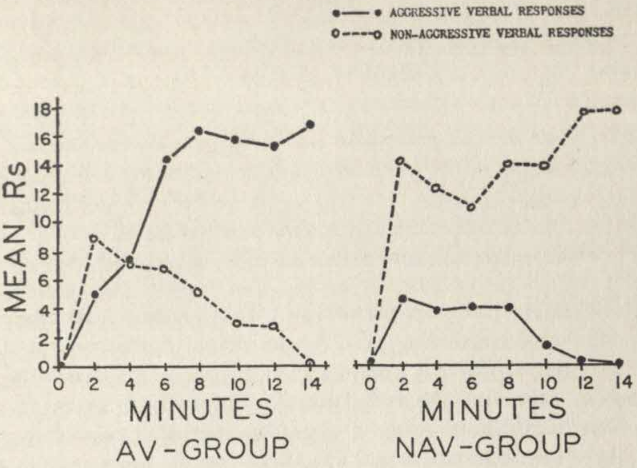

FIGURE 1—Mean number of aggressive and nonaggressive verbal responses over successive two-minute periods in the verbal conditioning period. Aggressive verbal conditioning group presented separately from the nonaggressive verbal conditioning group.

RESULTS AND DISCUSSION

Changes in the verbal behavior as a function of the verbal conditioning are presented in Figure 1. Number of aggressive and nonaggressive verbal responses was summed and averaged over successive two-minute periods for the various *S*s; the average aggressive and nonaggressive verbal responses for each group are presented in Figure 1. Since some *S*s reached the criterion of adequate verbal conditioning after six minutes, the last two-minute intervals give the verbal responding of decreasing numbers of *S*s. Thus, Figure 1 is only an approximate presentation of the changes in verbal behavior. The figure indicates that both objectives of the verbal conditioning were affected: first, there is an initial increase in rate of responding, and secondly, there is subsequent differentiation of aggressive and nonaggressive verbal responses. It should be noted that these *S*s emitted a very limited range of verbal responses. This was particularly true of the AV-group which in general was limited to "bad doll" or "dirty doll."

The effect of the verbal conditioning on the subsequent play behavior was calculated by taking the total number of responses for each *S* on the striking dolls and the ball-toy and converting this into the following score: 100 × total responses on the dolls ÷ total responses on dolls and ball. This percentage gives the relative amount of aggressive play behavior and was calculated for each *S* both before and after the verbal conditioning pro-

TABLE 1

MEAN PROPORTION OF AGGRESSIVE RESPONDING BEFORE AND AFTER VERBAL CONDITIONING

	Before Verbal Conditioning	*After Verbal Conditioning*
Aggressive conditioning	55.2	78.4
Nonaggressive conditioning	55.4	59.4

NOTE.—Proportion = 100 × total Rs on doll / total Rs on doll + ball.

cedure. The group means are presented in Table 1. As can be observed, the level of aggressive responding prior to the verbal conditioning is identical for the two groups. There is a significantly higher proportion of aggressive play behavior following the verbal conditioning period for the AV-group *Ss* who were conditioned to make aggressive verbal responses ($t = 2.326$, $df = 12$, $p < .05$, two-tailed test). In other words, some control of nonverbal aggressive behavior was achieved by manipulating the children's verbal aggressive behavior.

In evaluating these results, four possibilities should be considered. First of all, the aggressive verbal response, since it is reinforced and not punished, becomes a discriminative stimulus which marks the occasion when nonverbal aggressive behavior will be reinforced, or at least not punished. In popular terms, the aggressive verbal behavior provides a "green light" for subsequent aggressive nonverbal responding.

The second interpretation to be considered deals with the reinforcing stimuli held in common by the verbal aggressive and the nonverbal aggressive behavior. The common reinforcing stimuli may be primary (e.g., tension reduction) as well as secondary (e.g., other forms of self-stimulation). Frequently it appears that in everyday life verbal behavior achieves consequences similar to those of nonverbal behaviors (e.g., both may achieve the removal of some noxiously stimulating person); hence one would expect them also to have numerous secondary reinforcers in common. To the extent that the two response systems have reinforcing stimuli in common, it would be unlikely that an operation upon one system would not also change characteristics of responding in the other system. Both an increase and a decrease could be brought about in subsequent aggressive nonverbal behavior depending upon the operations on the verbal aggressive behavior. For example, one should be able to bring about a "sensitization" for aggressive reinforcers by presenting these briefly to the verbal mode of responding. This sensitization should bring about a subsequent increase in nonverbal responding for that reinforcer. This may well account for the effect observed in the present study. Insofar as the reinforcing stimuli are common between the two response systems, it also should be possible to "satiate" the organism on the reinforcer by presenting it amply for the one mode of responding, whereupon one would observe a subsequent decrease in responding for that reinforcer by the other mode of responding. The possibility of reducing nonverbal aggressive behavior by providing frequent and intensive verbal aggressive expression has been long recognized in clinical psychology, usually referred to by such terms as "cathartic expression." There is experimental evidence which supports this notion; for example, Feshbach (3) observed a decrease in aggressive behavior immediately following the writing of aggressive TAT stories.

The third manner in which verbal and nonverbal behavior may be seen to interact attributes to the verbal behavior a "directing" influence upon nonverbal behavior. This formulation presupposes that in the history of the person certain verbal and nonverbal behaviors have occurred together in such a manner that the verbal response becomes a discriminative stimulus for the nonverbal behavior.[3] This formulation may lead one to expect that one can vary the amount of generalization from verbal to nonverbal behavior by varying the degree to which the verbal behavior denotes the nonverbal behavior. For example, one would be led to expect more generalization than was observed in this study if the children in the AV-group had been reinforced for "doll should be spanked." This hypothesis is supported by observation of the behavior of some of the *S*s in the NAV-group. For example, one of these *S*s was reinforced for "what shall I say?" As this response increased in frequency with reinforcements, *S* got up from his chair and walked toward *A* to address his question to her.

Fourth and lastly, it is assumed that aggressive behavior has been associated with aversive stimuli in the past history of the child. It is conceivable that the occurrence of the verbal aggressive response allows for some extinction of the conditioned aversive stimuli associated with that response and that the effect of this extinction generalizes to the nonverbal response. If conditioned aversive stimuli suppress or inhibit aggressive responding, the effect to be expected would be an increase in subsequent nonverbal aggressive responding. In view of the resistance of aversive stimuli to extinction, this relationship is perhaps the least likely explanation of the present data.

SUMMARY

The purpose of the present research was to determine the effect of strengthening one class of responses on another. A bar-pressing response reinforced by aggressive doll action was observed immediately after the children had undergone a verbal conditioning session during which one group of children was reinforced for emitting aggressive verbal responses; the other group was reinforced for nonaggressive verbal responses. The

---

[3] Informal observation on the *S*s while they were being verbally conditioned further confirms that such behavior was being emitted; many *S*s in the AV-group raised their voices and became quite loud, thrashed around, pounded the table, etc.

results of the study gave evidence for an increase in aggressive nonverbal behavior following reinforcement of aggressive verbal behavior. In other words, some control of nonverbal aggressive behavior was achieved by manipulating verbal aggressive behavior.

In evaluating these results, four possibilities should be considered: (a) the verbal aggressive behavior becomes a discriminative stimulus which marks the occasion when nonverbal aggressive behavior will be reinforced or at least not punished. (b) To the extent that the two response systems have reinforcing stimuli in common, operating upon one system might also change characteristics of responding for these stimuli by the other system (e.g., by sensitization of or satiation for the common reinforcer). (c) The verbal response has a "directing" influence on the nonverbal response since it functions as a discriminative stimulus for that response. (d) Occurrence of the verbal aggressive response allows for some extinction of the conditioned aversive stimuli associated with that response; the effect of this extinction generalizes to the nonverbal response and thereby reduces the amount of aversive stimuli inhibiting the nonverbal aggressive responding.

REFERENCES

1. Bijou, S. W. Patterns of reinforcement and resistance to extinction in young children. *Child Develpm.*, 1957, 28, 47-54.
2. Esper, E. A. Language. In C. Murchison (Ed.), *A handbook of social psychology*. Clark Univer. Press, 1935. Pp. 417-460.
3. Feshbach, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, 50, 3-11.
4. Kurtz, K. H., & Hovland, C. I. The effect of verbalization during observation of stimulus objects upon accuracy of recognition and recall. *J. exp. Psychol.*, 1953, 45, 157-164.
5. Lehmann, A. Über Wiedererkennen. *Phil. Stud.*, 1888, 5, 96-156.
6. Lovaas, O. I. Effect of exposure to symbolic aggression on aggressive behavior. *Child Develpm.*, 1961, 32, 37-44.
7. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, 43, 173-176.

SIMULTANEOUS AND SUCCESSIVE DISCRIMINATION LEARNING IN CHILDREN[1,2]

LEWIS P. LIPSITT*

Brown University

In discrimination learning with infrahuman organisms, two procedures for the presentation of the stimuli to be discriminated are available. In a discrimination problem to be learned in a T-maze or a Lashley jumping apparatus and involving a black and a white stimulus, for example, the "simultaneous" procedure involves the presentation of a black and a white stimulus on each trial with position counterbalanced. The "successive" procedure entails presentation of only one of the discriminative stimuli on each trial. The reinforced responses in the simultaneous situation might be going left to black in a black-left, white-right arrangement and going right to black in a white-left, black-right arrangement. Approach to black, therefore, would consistently result in reinforcement. The reinforced responses in the successive situation might be going right to a black presentation and left to a white presentation.

Spence (7) has described simultaneous discrimination learning as resulting from the strengthening of response to the reinforced component of the stimulus complex until it attains sufficiently greater strength than competing incorrect response tendencies to result in consistently correct choices.

---

* Department of Psychology, Brown University, Providence 12, Rhode Island.

[1] Portions of this paper were presented in the symposium, "Recent Developments in Experimental Methods with Children," at the meetings of the Amer. Psychol. Ass., Washington, D. C., 1958.

[2] This research was supported initially by a Grant-in-Aid from the Sigma Xi - RESA Research Fund and subsequently by a National Science Foundation grant. The writer is indebted to Mr. Albert F. Ehnes, Superintendent of the Seekonk, Massachusetts, Public Schools, and to Mr. Edward Fitzgerald, Superintendent of the Bristol, Rhode Island, Public Schools for their cooperation and to the principals and teachers who helped make these studies possible.

Child Develpm., 1961, 32, 337-347.

Spence (8) has also said that the successive problem can be solved as a patterned discrimination, i.e., by *S*'s approaching certain patterns or combinations of reinforced stimulus components, while response to other patterns is not reinforced. He suggested that on this basis the successive problem should be more difficult than a simultaneous learning problem involving the same stimuli, because the stimulus compounds to be discriminated on each trial are more similar than in the simultaneous problem, and he presented evidence for this proposition from rat studies.

Not all infrahuman studies have supported the contention that simultaneous stimulus presentation produces faster learning than successive presentation. Several studies have suggested that at least under some experimental circumstances successive provides better performance than simultaneous stimulus presentation (1, 2, 9). In general, however, those experimental instances in which successive proved easier than simultaneous learning have involved a different kind of response than those characteristic of the T-maze or typical Lashley jumping situation. When the animal is required to "orient towards and approach the stimulus complex (path, door, alley, window) containing the positive stimulus cue (white, form, etc.) . . ." (8, p. 89), simultaneous stimulus presentation tends to result in better discrimination learning than successive. When the animal is required, however, to make a response to a locus removed from the stimulus source (1, 2, 9), it is not unusual for successive to prove either equal to or easier than the simultaneous problem. Furthermore, the relative difficulty of the simultaneous and successive problems has been reported to be influenced greatly by the similarity of the stimuli to be discriminated (1, 3, 4, 8).

The studies reported here investigate the relative discrimination learning performances of children under the simultaneous and successive stimulus presentation procedures, varying the nature of the response required and stimulus similarity.

EXPERIMENT 1[3]

Experiment 1 involves a learning situation for children in which the to-be-learned responses are to loci (buttons) removed from the stimulus sources (the windows or apertures).

Method

Apparatus.[4] The apparatus was a black plywood box, 12 in. high by 12½ in. deep. The front panels were 25¼ in. wide. There were three stimulus apertures each faced with a milk glass window on the upper panel and three response buttons on the lower panel. Distance from the three stimulus apertures to their respective response buttons was 6½ in. Behind each of the windows were five different colored bulbs (GE C-6 Series type) set in a metal reflecting unit. Three selector switches, one for each window, were situated at the back of the apparatus where *E* could preselect which color would appear in each window. Another selector switch enabled *E* to preselect which of the three buttons would be correct on each trial. A jeweled reflector light, located just above the middle button, would be activated when *S* pushed the correct button. A toggle switch activated the entire circuit so that (a) the preselected lights came on in the three windows and (b) response to the appropriate button would turn out the stimulus lights and activate the signal light, while wrong responses would merely turn out the stimulus lights. Thus, correction on the part of the *S* was not possible. The buttons were attached to 1/16 in. metal rods which passed through small holes in the response panel, each rod meeting a microswitch inside.

---

[3] This experiment was reported at the meetings of Eastern Psychol. Ass., Philadelphia, April, 1958.

[4] The apparatus, financed by the Ruth Holton Memorial Fund, was constructed while the writer was at the Iowa Child Welfare Research Station. Thanks are due Mr. John Peterson, who aided in its construction.

Subjects and Design. The *S*s were 40 fourth grade children obtained from the Bristol, Rhode Island, Public Schools and randomly assigned to one of four groups. Half of the total number of *S*s were administered a simultaneous discrimination learning problem, and half a successive discrimination problem. Half of the *S*s in each of these groups were run in a "similar-stimulus" condition, and half in a "distinctive-stimulus" condition. For the similar stimulus group, the colors were red, pink, and blue. For the distinctive stimulus group, the colors were red, green, and blue. On the assumption that red and pink are more similar physically than red and green, it was presumed that a greater amount of generalization would occur for the similar stimulus group, thus retarding learning involving these stimuli.

Procedure. A three-stimulus, three-response procedure was employed for all *S*s. Half of the *S*s were administered the simultaneous procedure, and half the successive stimulus procedure. The manner of stimulus presentation for the simultaneous and successive conditions closely follows procedures used in infrahuman investigations of two-stimulus problems. An *S* under the simultaneous condition was presented with the three stimuli exposed together, one in each aperture. Response to the button under one of these colors was consistently followed by the onset of the reward light, while response to either of the other buttons on that trial was not.

The order of stimulus presentation was the same for all *S*s under the simultaneous condition, and the successive order was the same for all *S*s in those groups. There were three stimulus arrangements (A: 1, 2, 3; B: 2, 3, 1; C: 3, 1, 2) possible for the simultaneous groups, and three (A: 1, 1, 1; B: 2, 2, 2; C: 3, 3, 3) for the successive groups. Each of these arrangements occurred in each block of three trials. These three arrangements may be ordered in six different ways (ABC, BAC, BCA, CBA, CAB, ACB). The order in which these six combinations were administered was determined randomly (with the restriction that identical stimulus presentations should not occur on successive trials) to provide a total of 18 trials. The 18-trial block was repeated to produce two cycles or a total of 36 trials, the maximum number administered.

Thus, for both kinds of discrimination, each response button is correct once within each three-trial block, and no position response is ever correct two trials in a row. Furthermore, since the correct-response order is the same for both, neither group is provided mnemonic cues for possible response-sequence learning that is not provided for the other.

All *S*s performed individually in a room away from the classroom. The same instructions were given all *S*s, regardless of the condition of stimulus presentation to which they were assigned. When seated at the apparatus, they were informed that they would be shown some lights and that when the lights came on they were to press one of the buttons that they thought correct. In the beginning they would not know which buttons were correct, but after a while they would learn which buttons to push and then would not make so many mistakes. They were told that the green signal light would go on when they were correct, but that that light would not go on when wrong.

*S*s were trained until they reached a criterion of nine consecutively correct responses or for a total of 36 trials. A trial is here defined in the conventional manner as the sequence proceeding with the onset of stimuli followed by the execution of one response, offset of the stimuli, and where appropriate onset of the reward light.

Results

The response measure here is the number of correct responses in 36 trials with correct responses assumed for those trials for any *S* beyond nine consecutively correct. Figure 1 compares the number of correct responses made over 36 trials by the simultaneous and successive groups, disregarding similarity of stimuli. While the two groups perform about equally and at chance level during the first block of six trials, the successive group overtakes the simultaneous and maintains its superiority through the remaining five blocks. Table 1 presents the over-all means and standard deviations for each of the four groups. An analysis of variance over the number-correct data, including analysis of trials effects, revealed that the superiority of successive over simultaneous training was reliable at the .01 level ($F = 7.39$, $df = 1, 36$). The main effect of stimulus similarity was not different from chance, but the trials by similarity interaction was reliable at the .05 level ($F = 4.04$, $df = 5, 180$), based on a tendency for the learning rate of the distinctive stimulus condition to be greater than that for the similar stimulus condition. Other than a trials effect significant at the .001 level, the only other reliable effect was the triple interaction ($F = 6.59$, $df = 5, 180$, $p = .001$) based on a tendency for there to be no differences in learning rate for the two successive groups, but a greater learning rate for the simultaneous-distinctive than the simultaneous-similar condition.

The major finding of this experiment is that under a procedure involving response to buttons removed from the stimulus source, successive stimulus presentation provided reliably better discrimination learning than simultaneous presentation.

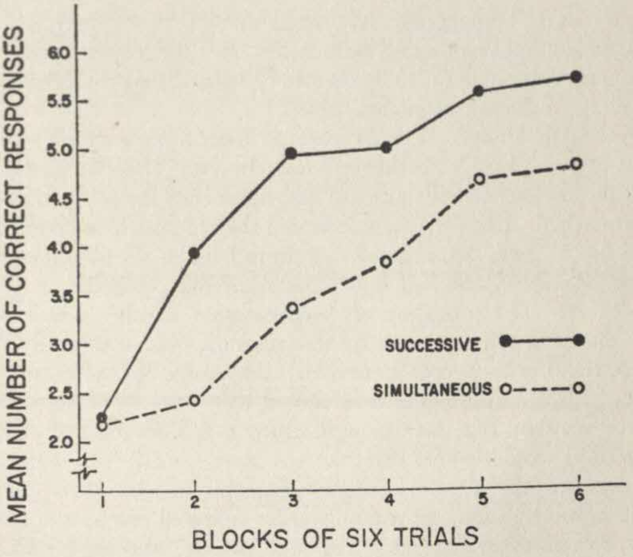

FIGURE 1—Learning curves for the simultaneous and successive groups of experiment 1, involving response to a locus removed from the stimulus source.

TABLE I

MEANS AND STANDARD DEVIATIONS FOR THE FOUR GROUPS OF EXPERIMENT 1

	Type of Stimulus Presentation				
	Simultaneous		*Successive*		*Total*
	M	SD	M	SD	M
Similar (red, pink, blue)	21.2	7.7	27.3	6.1	24.3
Distinctive (red, green, blue)	21.3	7.8	27.5	4.8	24.4
Total	21.3		27.4		

EXPERIMENT 2

This experiment was designed to determine the importance of the button-pushing response involved in experiment 1. In the present experiment, all conditions were the same as in the first experiment except that now *S*s were required only to point toward and touch the correct window on each trial.

Method

Apparatus. The apparatus was the same as that of experiment 1. The buttons, which were removable, were taken from the response panel leaving in their place very small (1/16 in. diameter) holes through which the button rods leading to the microswitches passed.

Subjects and Design. The *S*s were 20 fourth grade children from the Pleasant Street School in Seekonk, Massachusetts. These *S*s had not been subjects in previous experiments and had never seen the apparatus with the buttons attached. All *S*s were administered the red-pink-blue discrimination. Half of the *S*s were adminstered the stimuli under the simultaneous presentation procedure, and half under the successive procedure.

Procedure. The procedure of discrimination learning was exactly the same as that of experiment 1, except that response buttons were not available to *S*s. Instead, the *S*s were instructed that they would be shown some lights and that each time the lights were shown they were to point toward and touch the window that they thought correct. If they touched the correct window, they would be told that that was correct, while if they touched the wrong window the *E* would indicate that this was incorrect. Manner of stimulus administration and stimulus order followed exactly the same plan as that of experiment 1. *E* said "Yes" or "Wrong" after each trial.

Results

The learning curves for the simultaneous and successive discrimination learning groups are seen in Figure 2. Here it may be seen that the simultaneous stimulus procedure is clearly superior to that of the successive group throughout training. The simultaneous group started at chance level within the first three trials of the experiment and by the third block of six trials was performing 100 per cent correctly. The successive presentation group, on the other hand, did not as a group depart from chance performance through the 36 training trials, and only one subject of 10 solved the successive problem by the end of training. A median test of the over-all difference in performance between the groups resulted in a chi square with Yates' correction of 9.8 ($df = 1$), which is reliable beyond the .01 level of significance.

It seems quite clear from experiment 2 that, when the children are required to respond directly to the stimulus source, the simultaneous stimulus presentation results in better learning, in contrast to the finding of experiment 1 where *S*s responded to a locus some distance removed from the stimulus source.

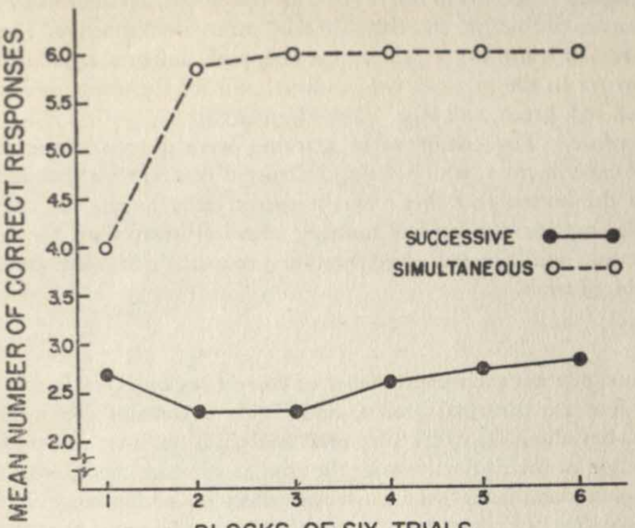

FIGURE 2—Learning curves for the simultaneous and successive groups of experiment 2, involving response directly to the stimulus source.

EXPERIMENT 3

The next experiment returns to the light-button procedure and investigates the relative ease of discrimination learning under the simultaneous and successive stimulus presentation procedures under two levels of stimulus similarity. The purpose of the present experiment was to test the limit of the successive superiority over simultaneous of experiment 1 by using a condition of extremely high stimulus similarity. It was thought that, at some high level of stimulus similarity, the simultaneous condition must produce better discrimination than the successive even under the light-button procedure, on the assumption that the psychophysical difference thresholds must be smaller for paired comparisons procedures than for methods of single stimuli. Thus, if the stimuli are made highly similar, the problem should be more soluble for simultaneous *S*s than for successive *S*s.

Method

Apparatus. The apparatus was the same as that of experiments 1 and 2, with the buttons replaced.

Subjects and Design. The *S*s were 40 fourth grade children from the Newman Avenue and North Schools in Seekonk, Massachusetts, who had not previously participated in discrimination learning experiments. They were assigned randomly to one of the four conditions: simultaneous-similar, simultaneous-distinctive, successive-similar, successive-distinctive. The colors for the similar stimulus conditions were red, pink, and orange (more similar than any set in the previous two studies), and for the distinctive stimulus condition red, green, and blue.

Procedure. The conditions of learning were essentially the same as those of experiment 1, with *S*s being instructed that on each trial they were to push the button that they thought correct after looking at the lights. Particular caution was used in insuring identical instructions for both the simultaneous and successive presentation groups. All *S*s were trained for a total of 36 trials.

Results

Figure 3 shows the mean number of correct responses made in 36 trials by the four experimental groups, and Table 2 contains the means and standard deviations on which a factorial analysis of variance was performed. Comparison of the distinctive with the similar stimulus conditions indicates that stimulus similarity has an over-all effect on performance significant at the .001 level ($F = 18.82$, $df = 1, 36$). It may be noted that the mean performance of the distinctive-successive group is somewhat superior to that of the distinctive-simultaneous group (as in experiment 1), but for the highly similar stimuli simultaneous presentation provides clearly superior performance to successive. This interaction between type of discrimination and level of stimulus similarity fell just short of significance at the .05 level ($F = 3.92$, $df = 1, 36$), and the difference between simultaneous and successive means under the similar-stimulus condition yielded a t of 2.33 ($df = 18$, $p = .05$).

While this experiment did not reliably reproduce the successive over simultaneous superiority of experiment 1, the direction of the difference is the same, and the expected inversion of the simultaneous and successive means with a great increase in stimulus similarity is found. Thus, it appears

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR THE FOUR GROUPS OF EXPERIMENT 3

	TYPE OF STIMULUS PRESENTATION				
	Simultaneous		*Successive*		*Total*
	M	SD	M	SD	M
Similar (red, pink, orange)	22.9	7.6	15.9	4.9	19.4
Distinctive (red, green, blue)	27.9	7.2	29.3	5.3	28.6
Total	25.4		22.6		

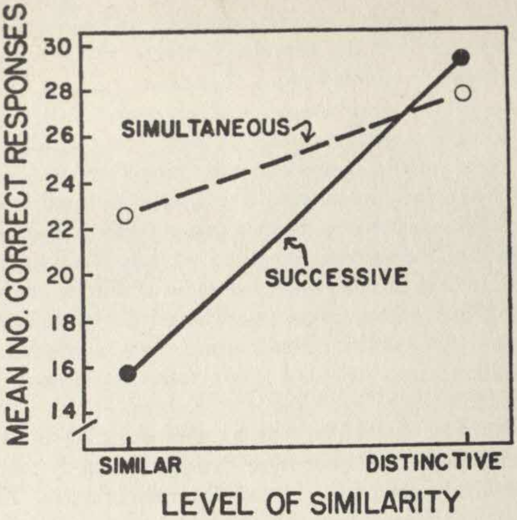

Figure 3—Comparisons of mean number of total correct responses from experiment 3 for the simultaneous and successive groups under two levels of stimulus similarity.

that for children both stimulus similarity and the nature of the response required are pertinent parameters in the determination of the relative ease of learning simultaneous and successive discrimination problems.

Discussion

The three studies reported above considered together support the propositions that in children (a) simultaneous stimulus presentation tends to provide better visual discrimination learning than successive when the response required is to the source of stimulation, but that, (b) if the response required is to a locus removed from the stimulus source, either no differences occur or an inversion results with successive presentation yielding the better discrimination performance; furthermore, (c) this relationship is affected by degree of stimulus similarity, such that under conditions of high similarity, simultaneous presentation provides better learning than successive while with highly distinctive stimuli this difference may disappear or result in an inversion with successive superior to simultaneous. Thus, conditions which would tend to maximize simultaneous superiority over successive would involve high stimulus similarity and response to the stimulus source, with the obverse conditions tending to yield successive superiority over simultaneous.

Several studies with infrahuman organisms (1, 2, 4, 8, 9) have found some of the above phenomena, but few studies have investigated the problem of simultaneous and successive stimulus presentation in humans and fewer still have investigated the interactions of parameters likely to lead to differential learning of the simultaneous and successive problems. Loess and Duncan (3) have shown that, under a highly similar stimulus condition or difficult discrimination, simultaneous proved better than successive learning, but when the stimuli were less difficult to discriminate there was no statistical difference between the two procedures and in fact there tended to be superiority of the successive method. Perkins, Banks, and Calvin (6) investigated the effect of delay on learning in children. Their situation involved touching and picking up the discriminative stimuli, and simultaneous presentation tended to provide better learning than successive, with no effect whatever resulting from delay (a combination of delayed reaction and delayed reward).

Recently, Murphy and Miller (5) have reported on a study with children using an experimental procedure designed to study the generality of a finding previously obtained with infrahuman primates. Their stimulus presentation procedures involved only the simultaneous method. They found that it is much easier for a child to learn a discrimination problem when he is responding directly to the stimulus source and that the problem becomes very difficult if the task is changed so that the child must now respond to a locus removed from the stimulus source, even if the child has had previous training on the same discrimination problem under the approach-to-stimulus procedure. The results of the present studies would suggest that such results would be most likely in simultaneous discrimination problems and that this phenomenon may not prevail so strikingly under a successive stimulus presentation procedure.

The present experiments suggest that the stimulus similarity continuum warrants more definitive study as it affects simultaneous and successive discrimination in children. They further suggest that other parameters (such as distance from the stimulus source of the response locus) could possibly produce similar behavioral effects and be superimposed upon the stimulus similarity dimension. It is possible also that other experimental manipulations which tend to depress discrimination learning generally, such as increasing delay of reward, decreasing age, and large spatial or temporal separation of the stimuli, might have similar effects on performance in the simultaneous and successive problems. Stimulus generalization may be the pertinent mechanism producing the inversion effects of the present studies and could result in similar findings when produced through other procedures.

SUMMARY AND CONCLUSIONS

The results of three discrimination learning studies of normal fourth grade children are reported. Comparisons of the methods of simultaneous and successive stimulus presentation are made under different levels of stimulus similarity and two types of response, either directly to the stimulus source or to a locus (buttons) removed from the stimulus source. The results of the three studies agree very well with the following propositions:

1. When response is directly to the stimulus source, simultaneous stimulus presentation tends to result in better learning than successive.

2. When the response is to a locus removed from the stimulus source, successive stimulus presentation may result in performance equal to or better than simultaneous.

3. The nature of the response involved in discrimination learning interacts with stimulus similarity, such that in procedures involving response to a locus removed from the stimulus source simultaneous may produce better learning than successive if the stimuli are very highly similar.

It is concluded that the relative ease of simultaneous and successive discrimination learning in children is a joint function of stimulus similarity and the type of response required. It is suggested further that other manipulable conditions undoubtedly interact with the two dealt with here and that possibly stimulus generalization is the mechanism underlying the effects of each.

REFERENCES

1. BITTERMAN, M. E., TYLER, D. W., & ELAM, C. B. Simultaneous and successive discrimination under identical stimulation conditions. *Amer. J. Psychol.,* 1955, 68, 237-248.
2. BITTERMAN, M. E., & WODINSKY, J. Simultaneous and successive discrimination. *Psychol. Rev.,* 1953, 60, 371-376.
3. LOESS, H. B., & DUNCAN, C. P. Human discrimination learning with simultaneous and successive presentation of stimuli. *J. exp. Psychol.,* 1952, 44, 215-221.
4. MACCASLIN, E. F. Successive and simultaneous discrimination as a function of stimulus similarity. *Amer. J. Psychol.,* 1954, 67, 308-314.
5. MURPHY, J. V., & MILLER, R. E. Spatial contiguity of cue, reward, and response in discrimination learning by children. *J. exp. Psychol.,* 1959, 58, 485-489.
6. PERKINS, M. J., BANKS, H. P., & CALVIN, A. D. The effect of delay on simultaneous and successive discrimination in children. *J. exp. Psychol.,* 1954, 48, 416-418.
7. SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.,* 1936, 43, 427-449.
8. SPENCE, K. W. The nature of the response in discrimination learning. *Psychol. Rev.,* 1952, 59, 89-93.
9. WEISE, P., & BITTERMAN, M. E. Response selection in discrimination learning. *Psychol. Rev.,* 1951, 58, 185-195.

SECONDARY REINFORCEMENT IN A DISCRIMINATION PROBLEM WITH CHILDREN[1,2]

ALAN H. LEIMAN, JEROME L. MYERS,
and NANCY A. MYERS*
University of Massachusetts

There have been relatively few studies of the role of secondary reinforcers in motivating the performance of children. Nor are many data available on the effectiveness of secondary reinforcers in maintaining discrimination learning. The present study was designed to contribute information on both of these points, by investigating the effectiveness of secondary reinforcers in discrimination learning in children.

Two relevant studies have been carried out with adults. Hubbard (4) trained adults with a red light as a primary reinforcer and an accompanying tone as a secondary reinforcer for correct discriminative responses. He obtained a highly significant secondary reinforcement effect; the experimental group made more correct responses during extinction trials than did either of two control groups. Fort, Myers, and Myers (3) also investigated discrimination problem performance, with a more clearly defined primary reward than that used by Hubbard. Hubbard's "primary reward" was a red light used to indicate correct responses; Fort and Meyers gave *Ss* a poker chip, exchangeable for five cents at the end of the session. The results of the two studies were highly similar.

---

* Department of Psychology, University of Massachusetts, Amherst.

[1] This research was supported by funds from an NIMH Grant, M-2620.

[2] The authors wish to thank Florence Leiman for her assistance in collecting the data.

Child Develpm., 1961, 32, 349-353.

Utilizing the design of these two investigations, the present experiment attempted to determine whether the findings of Hubbard and those of Fort and Myers could be extended to grade school children. Appropriate modifications in the complexity of the task and in the primary reward were made.

METHOD

Subjects

The *S*s were 45 fifth grade children of Cold Springs School in Belchertown, Massachusetts.[3] They were assigned to three experimental groups of 15 each. An attempt was made to balance groups for age and for sex. Two children stopped before acquisition was completed (one from the C1 and one from the C2 group), and three in the C2 group did not reach criterion. These children were replaced by others. Once reaching criterion during acquisition, all of the children stayed for the extinction period. No special procedures were used for keeping the children in the extinction period.

Apparatus

The apparatus consisted of the *S*'s light panel box and the *E*'s control box. Mounted on the light box was a row of three identical red lights. A switch was mounted below each of the three lights. One, and only one, light was on for each trial. The order in which the lights came on was automatically determined by a stepping relay wired to produce a random series of 44 lights, with the one restriction that no light occurred more than three times in succession. When *S* pushed the switch directly below the light which was on during a particular trial, an M & M candy was dispensed to a chute, which delivered it to an opening cut in the panel box. *S*'s response also lit a corresponding light on *E*'s control box, and *E* recorded which switch was pressed by the *S* on each trial. The *E* also had easy access to switches on the control box which operated candy and buzzer mechanisms.

Procedure

Each *S* was seated in front of the light panel box, and *E* explained the rules of the game: whenever one of the lights came on, he was to press one and only one switch. If the switch pressed was the correct one for that light, a candy would be delivered. *E* then pressed one incorrect switch to demonstrate what might happen. *S*s were also told they could leave at any time, by indicating that they wished to stop the game. *E* then took a position behind the apparatus so that minimum interaction with *S* took place during the experiment.

---

[3] The authors wish to thank Carl J. Peterson, Superintendent of Schools in Belchertown, Massachusetts, and Mr. Donald Geer, Principal of the Cold Springs School, for their co-operation in permitting us access to the students and the school facilities.

The *S*s were divided into three groups, a secondary reinforcement group (S^r) and two control groups (C1 and C2). Training for all groups continued until *S*s scored 10 consecutive correct responses. Children who elected to leave before this time or who did not reach criterion by the end of the third presentation of the cycle (132 trials) were replaced by another child of the same sex and similar age. Extinction began on the trial immediately following the tenth consecutively correct training response and continued for 75 trials. The S^r group received a buzz paired with candy during training and the buzzer alone during extinction for correct responses. Group C1 was also trained with a buzz and candy, but received neither the candy nor buzz during extinction. Group C2 received the candy alone for correct responses during training and the buzz alone for correct responses during extinction.

RESULTS

Acquisition data were analyzed to determine whether any differences existed between groups prior to extinction. The groups did not differ significantly on number of trials for reaching criterion (Kruskal-Wallis test, $H = .30$, $p < .5$). This strongly indicates that all groups entered the extinction phase without meaningful differences.

The percentage of correct responses for blocks of 15 extinction trials each are presented in Figure 1. The S^r group (buzz and candy in training and buzz in extinction) made many more correct responses during extinc-

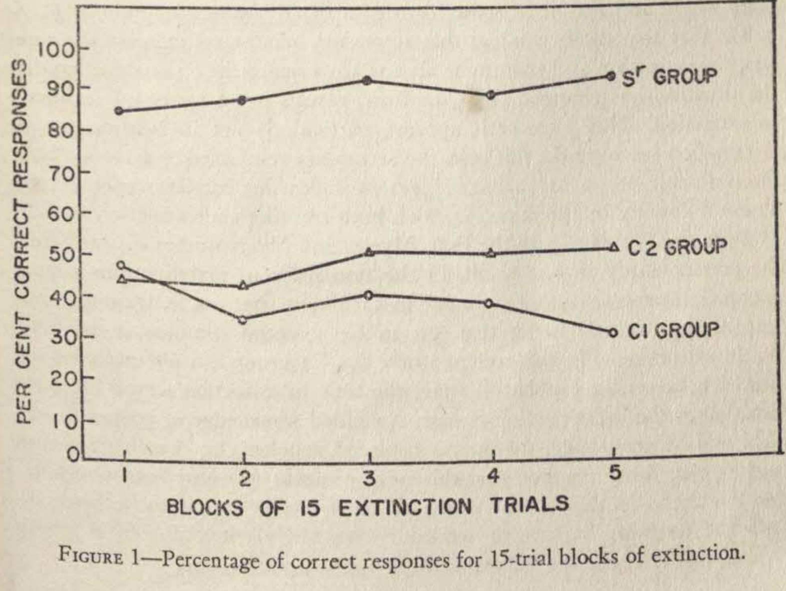

FIGURE 1—Percentage of correct responses for 15-trial blocks of extinction.

tion than did the control groups. The mean percentage of correct responses for the S^r group was 89.4. Group C2 (candy during training, buzz during extinction) made 48.3 per cent correct responses. Fewest correct responses (38.8 per cent) were made by group C1 (candy and buzz during training, but neither during extinction). These *S*s were apparently operating very close to a chance level (33.3 per cent).

An analysis of variance was performed on the arcsin transformations of these data and yielded a significant difference between the groups ($F = 26.38$, $p < .01$). Trials and the Groups-by-Trials interaction were not significant.

Duncan's Multiple Range test (2) was applied to compare each of the groups with one another. Group S^r was significantly different from each of the control groups ($p < .01$). The control groups did not differ significantly from each other.

DISCUSSION

The present study was an attempt to add to the data on the role of secondary reinforcement in human performance, by considering its effectiveness in a discrimination problem, using children as *S*s. It is apparent that a strong secondary reinforcer can be established in children. The secondary reinforcer appears to be equally effective in children and adults, since the percentages of correct responses are similar for the S^r groups in three experiments, the present one, that of Hubbard (4), and that of Fort, Myers, and Myers (3). The most parsimonious explanation for this consistently high performance of secondary rewarded *S*s in discrimination tasks would seem to lie in some form of a discrimination hypothesis (1, 5, 7, 8). This hypothesis assumes that secondary reinforcers increase the similarity of extinction and training trials and thus retard the *S*'s recognition that the situation has changed. This, in turn, results in an increased resistance to extinction. This hypothesis appears particularly apt in relation to discrimination learning. In this case, the secondary reinforced *S*s have a clearly discriminable stimulus (a buzz) present following correct responses and absent following incorrect responses on both training and extinction trials.

Both the Hubbard and the Fort, Myers, and Myers studies differed from the present study in one result. In the two previous investigations control group C2 (neutral stimulus present in extinction, but not in training) performed significantly better than group C1 (neutral stimulus in training, not in extinction). In the present study the C2 group did not perform significantly better. In Hubbard's study the tone in extinction served the same function as the light previously had; it yielded knowledge of results. In the Fort and Meyers study, the buzzer replaced a poker chip worth five cents, and *S*s may have assumed that this new stimulus was also "exchangeable" for five cents. In the present study, no such factors appear to be involved, and the resulting picture of secondary reinforcement and control group effects would therefore appear to be a more accurate one.

SUMMARY

The present study investigated the role of a secondary reinforcer in a discrimination problem for children and showed that a secondary reinforcer can be established with children, as has already been seen in studies with adults. Three experimental groups were used. The secondary reinforcement group (S^r) received candy and buzz for correct responses during training and a buzz during extinction. This group made significantly more correct responses during extinction than did either of two control groups. The first control group ($C1$) received candy and a buzz for correct responses during training, but neither during extinction. The second control group ($C2$) received candy during training and a buzz during extinction for correct responses. The number of correct responses during extinction for group $C2$ was less than that of group S^r and slightly more than that of group $C1$.

REFERENCES

1. BITTERMAN, M. E., FEDDERSON, W. E., & TYLER, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, 66, 456-464.
2. DUNCAN, D. B. A significance test for differences between ranked treatments in an analysis of variance. *Virginia J. Sci.*, 1951, 2, 171-189.
3. FORT, J. G., MYERS, J. L., & MYERS, N. A. Secondary reinforcement in a discrimination problem. *J. gen. Psychol.*, in press.
4. HUBBARD, W. R. Secondary reinforcement of a simple discrimination in human beings. *J. exp. Psychol.*, 1951, 41, 233-241.
5. MELCHING, W. H. The acquired reward value of an intermittently presented neutral stimulus. *J. comp. physiol. Psychol.*, 1954, 47, 370-373.
6. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* Wiley, 1951. Pp. 435-472.
7. MYERS, J. L. Secondary reinforcement: a review of recent experimentation. *Psychol. Bull.*, 1958, 55, 284-301.
8. ZIMMERMAN, D. W. Durable secondary reinforcement: method and theory. *Psychol. Rev.*, 1957, 64, 373-383.

PERSONALITY FACTORS RELATED TO JUVENILE DELINQUENCY[1,2]

DONALD R. PETERSON*
University of Illinois
HERBERT C. QUAY
Vanderbilt University
and THEODORE L. TIFFANY
Lincoln College, Illinois

Research on juvenile delinquency has often been handicapped by the behavioral heterogeneity of the delinquent population. Subjects have ordinarily had one common property, the commission of some legally defined delinquent act, and nothing more. It is little wonder that knowledge about cause and consequence has accumulated so slowly; the behavioral unity required for adequate scientific investigation is lacking.

One major approach to conceptual and methodological refinement has involved the delineation of typological and dimensional constructs of various kinds. Before examining social correlates of delinquency, Reiss (17) classified subjects into three groups, those with relatively weak ego controls, those with defective superego controls, and those with relatively integrated personal controls. Similarly, Jenkins and Glickman (7) have offered a basis for classifying delinquents into an unsocialized aggressive group, a socialized delinquent group, and an emotionally disturbed group. Both these efforts, and others like them, represent a considerable advance over earlier attempts to study juvenile delinquency without further specification, but their value is limited in various ways. Most of them are stated in typological, rather than dimensional, terms, and there is some reason to doubt the very existence of nonarbitrary personality types among juvenile delinquents (18). Even more fundamentally, the earlier formulations are typically based on case history information, which is necessarily indirect, notoriously vulnerable to subjective bias, and inclined to be unreliable.

A second approach to methodological refinement has led to the construction of a number of tests which distinguish delinquents from non-

---

* Department of Psychology, University of Illinois, Urbana.

[1] This study was supported by a grant from the University Research Board, University of Illinois.

[2] So many people have helped in the conduct of this research that we cannot possibly thank them all. We are especially grateful, however, to Earl K. Dryden, Chief Psychologist at the Illinois State Training School for Boys, St. Charles; E. B. Daly, Principal of Thomas Kelly High School in Chicago; C. S. Prigmore, Superintendent, and R. W. Edwards, Psychologist at the Tennessee State Vocational Training School for Boys; and to A. D. Hancock, Principal of Cumberland High School in Nashville, Tennessee, for their help in gathering data for the investigation.

Child Develpm., 1961, 32, 355-372.

delinquents on a behavioral rather than a legal basis (4, 9, 11, 14). Some of these instruments have proved remarkably effective in accomplishing the tasks for which they were designed (10, 12), but they fail to provide any improvement in conceptual unity. The items comprising the scales are nearly as heterogeneous in meaning as delinquents are in behavior.

Various difficulties with the above approaches can be overcome if data are restricted to behavior observations or products of some known relevance to delinquent activity and if appropriate analytic techniques are employed to define any structural concepts inherent in the data. Many kinds of behavior meet the first qualification. For this study, we have chosen questionnaire responses of demonstrated effectiveness in differentiating delinquents from nondelinquents. A number of statistical procedures have been proposed for the elaboration of structural concepts. We have elected to use centroid factor analysis.

In a previous study (13), two questionnaires which distinguished reliably between delinquents and nondelinquents were administered to 116 inmates of a boys' training school and to 115 male high school students. Responses were factorized, and five reasonably unitary, comprehensible dimensions emerged. The first implied a tendency toward impulsiveness, a lack of emotional involvement with others, and a tough, rebellious, distrustful attitude toward authority. It was called "Psychopathy." The second factor also contained an element of impulsivity, but this was associated with tension, discouragement, guilt, and depression. The factor was labelled "Neuroticism." The remaining three factors accounted for relatively small proportions of variance. One of them implied a sense of failure and incompetence and was called "Inadequacy." The other two were interpreted as background factors relating to family dissension and a history of scholastic maladjustment.

Extension of the meaning and validity of the three personality factors was sought by investigating associations with other logically related variables (15). For a sample of delinquents, Psychopathy was significantly related to recidivism, length of institutionalization, the occurrence of problem behavior and disciplinary measures during incarceration, and a history of crimes against person rather than against property. Among the variables examined, Neuroticism was correlated reliably only with a measure of manifest anxiety (1). The Inadequacy factor was related negatively to Army Beta IQ and positively to length of institutionalization. While the correlations were generally of low magnitude, they were in nearly perfect accord with factor interpretations originally developed solely on the basis of item content.

The results of this previous research appeared sufficiently promising to encourage initiation of the present study. Our purpose is the same as before: to generate a set of unitary, independent, possibly meaningful personality constructs related to delinquent behavior in the belief that these will offer a more rigorously defined frame of reference and more powerful hypotheses for further research than are now available to investigators of juvenile delinquency. This study differs from the previous one in that the range of behavior has been greatly expanded, and the sample of subjects approximately doubled.

SUBJECTS

Four hundred six subjects (*S*s) participated in the investigation. Of these, 93 were in a state training school in Illinois, 93 were attending a Chicago high school, 110 were located in a state training school in Tennessee, and 110 were students at a high school in Nashville, Tennessee.

In selecting *S*s, an effort was made to maximize variance along the gross dimension of major concern, i.e., delinquency proneness, and to reduce variance to a minimum along various other dimensions which might be related to delinquent behavior. Thus, both delinquents and nondelinquents were examined, but an attempt was made to equate the groups in respect to several other characteristics. All *S*s were white, male, and in the same age range (10 to 18 years). Boys at the Illinois State Training School had all been residents of Chicago, and the residential area for Chicago high school students resembled that of the "average" delinquent in regard to social and cultural factors tending to foster delinquency. Comparability of the latter kind was achieved by locating the home address of each Chicago delinquent, tabulating delinquency rate (19) for the part of town in which he lived, and obtaining the mean of those rates. Choice of a high school was then determined by its location; it was situated in an area with the same delinquency rate as the mean for adjudged delinquents in the sample. The relatively small size of the population in the Tennessee training school, the diversity of background of the inmates, and the lack of information on delinquency rate for home communities prohibited similar rigor in selecting *S*s from Tennessee. An attempt to achieve at least urban-rural comparability was maintained, however, by selecting nondelinquent *S*s from a high school on the outskirts of Nashville which drew most of its students from the city, but some of them from outlying districts. A majority of the Tennessee delinquents (65) were from the four largest cities in the state, and the remainder (45) were from smaller communities.

PROCEDURES

Questionnaires

A battery of four questionnaires, providing nearly exhaustive coverage of available items which differentiate between delinquent and nondelinquent boys, was administered to all *S*s.

PNI Attitude Schedule (PNI). The first questionnaire, containing 100 items, was called the PNI Attitude Schedule in reference to factors derived in previous research. Fifty-nine of the items had shown the requisite discriminatory power in earlier comparisons (4, 14) and appeared as salient variables marking Psychopathy, Neuroticism, and Inadequacy as defined in the factor analysis mentioned above (13). To secure better representation of the putative Inadequacy factor, 41 new items were written to express various facets of the dimension as the investigators conceived it. The Psychopathy (P), Neuroticism (N), and Inadequacy (I) items were systematically arranged in the order PINIPINI, etc., and the entire 100 items administered to the *S*s.

Only 68 statements, however, were included in the analysis. Since one criterion for selection of items was their ability to distinguish delinquents from nondelinquents, it was necessary to test the discriminatory power of the 41 new items and to cross-validate the results. One hundred fourteen delinquents from the Tennessee training school were compared with an equal number of nondelinquents from Tennessee as to performance on these items. Frequencies of endorsement were tabulated, chi square tests performed, and nine items selected as discriminating "significantly" between the two groups of *S*s. At this point, it became necessary to begin further analyses, and the nine items were retained in all subsequent work. On cross-validational comparison of 93 delinquents and 93 nondelinquents from Illinois, unfortunately, the nine-item subtest did not function effectively, and the 68 PNI items which were later factorized can only be described as follows: 25 of them discriminated between delinquents and nondelinquents and loaded on Psychopathy in the previous factorization; 25 differentiated between delinquents and nondelinquents and loaded on the Neuroticism factor; nine showed similar discriminatory power and loaded on Inadequacy; nine items, introduced to fortify measurement of Inadequacy, appeared at first to differentiate between delinquents and nondelinquents, but failed to do so on cross-validation.

The Personal Index (PI). The Personal Index by Loofbourow and Keys (11) was chosen as a second instrument for investigation. The test is fairly old, but we have assumed that the attitudes and other tendencies associated with delinquency have not changed radically since the discriminatory ability of the test was demonstrated, and it has been included, without modification, in the study.

The Personal Opinion Inventory (POI). A third questionnaire was assembled from certain items in the Minnesota Multiphasic Personality Inventory (MMPI), and a set of items developed in rehabilitation research in the U. S. Navy (5). The MMPI items, 33 in number, not only discriminated between previously classified delinquents and nondelinquents, but were, in addition, valid predictors of delinquent activity (6).

Items written for the Navy research were "designed to uncover individual susceptibility toward delinquent behavior in the military service" (5, p. 2). From a larger pool, 121 items were selected for their effectiveness in distinguishing Navy and Marine Corps enlisted men serving sentences in Naval retraining commands from enlisted men in various duty stations. The populations involved in this research are thus somewhat different from the delinquents and nondelinquents employed in the present study, but a considerable amount of research indicates that questionnaire items which differentiate between legally and socially defined groups of offenders and nonoffenders have a good deal of discriminatory generality, however the offenses are defined and regardless of subject age range. The Socialization Scale of the California Psychological Inventory, for example, has differentiated not only between delinquent and nondelinquent adolescents, but between guard house prisoners and duty soldiers in the Army, adult reformatory inmates and high school students, judged behavior problems and "good citizens" in school, and it relates rather impressively to assumed socialization over some 20 groups to which the scale has been applied (3, 4, 12). We cannot claim that the items defined by research on Navy and Marine Corps personnel actually differentiate between delinquent and nondelinquent adolescents, but there is good empirical reason to assume that the items would perform the desired discrimination, and this was our reason for including them in the battery.

The tendency among questionnaire authors to use items from previously developed instruments led to some overlap among tests. Seven of the 33 MMPI items also appeared in the PNI Attitude Schedule and were eliminated from the set now under discussion. Thirty-eight of the 119 items in the Navy questionnaire appeared either in the PNI or the MMPI, and these were also eliminated. Two more statements were omitted because they made reference to affairs of a specifically military kind. The remaining 105 items, 26 from the MMPI and 79 from the Navy scale, were entitled the Personal Opinion Inventory (POI) and were used as one test in this study.

The K D Proneness Scale (KD). The K D Proneness Scale by Kvaraceus (9, 10) was employed as a fourth questionnaire. It consists of 75 multiple-choice items which permit differentiation of delinquent and nondelinquent groups. The following item illustrates the format in which statements are presented: "Failure in school is usually due to (a) bad companions, (b) lack of ability, (c) lack of hard work, (d) unfriendly teachers." Since the response alternatives represent discrete choices, rather than points on continuua, the separate alternatives were employed as elemental variables in conducting the factorization. Of the 300 variables which thus appear in the scale (75 items × four alternatives) Kvaraceus has claimed discriminatory power for only 135, but the latter number was still too large to permit analysis, and the number of variables had to be reduced even further.

Elimination was accomplished by executing a preliminary factorization of some of the items, retaining those with the highest communalities, and pooling the latter with the remaining variables for a second analysis. The first 106 alternatives which perform the necessary differentiation were intercorrelated, and 10 centroid factors extracted. Eighty variables with communalities over .25 were then combined with the 29 remaining alternatives in the scale for the final factor analysis.

Data Analysis

The following analytic steps were performed with the aid of an electronic computer. First, phi coefficients of intercorrelation were computed for each of the four questionnaires, and 10 centroid factors were extracted from each correlation matrix. Decisions as to how many factors to rotate were largely arbitrary, although plots of variance removed by successive factors generally reinforced our decisions. Three factors were retained in the PNI Attitude Schedule, as directed by the hypothesis that Psychopathy, Neuroticism, and Inadequacy would emerge as factors in the instrument. Four factors were retained in each of the remaining three questionnaires. The four sets of factors were then rotated in accordance with Kaiser's varimax criterion (8).

The first order factors were given independent interpretation, as discussed below, and further analyses focused on relations among them. To reduce spurious intercorrelations between factors in the same test, factor scores were based on relatively pure items, viz., those with loadings $\geq .30$ on the factor in question and $< .20$ on all other factors. No differential weights were employed; scores were derived by algebraic addition, and the correlations among them determined.

As a final integrative analysis, second order questionnaire factors were defined. To permit first order factors statistical freedom to emerge at the second order, two scores had to be computed for each of the first order factors examined in the analysis. One vector in the Personal Index was eliminated because only three variables met our criteria for inclusion as salients, but two scores were obtained for each of the remaining 14 dimensions. The salient variables in each factor were rearranged in the order in which they had originally been presented to *S*s. One score was then derived from the first half of the items, and another from the second half of the items. Product-moment intercorrelations were calculated for the resulting 28 variables, 10 centroid factors were extracted, and three of the latter were rotated to orthogonal simple structure as defined by the varimax procedure.

RESULTS

All questionnaires used in the investigation, all centroid and rotated factor matrices, lists of salient variables complete with loadings and item numbers, and various tables pertinent to the second order analysis have been reported elsewhere.[3] Only the salient variables representing each factor will be discussed below, beginning with the first order dimensions derived from the questionnaires. In accordance with the usual convention, omission of sign before any item means that the response "true" is associated with the positive pole of the factor; a negative sign in parentheses at the beginning of an item indicates that the response "false" is associated with the positive pole of the factor.

---

[3] The four questionnaires, centroid factor matrices and rotated factor matrices pertaining to the analyses, complete lists of salient variables, and intercorrelations among first order factors have all been deposited as Document number 6606 with the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. A copy may be secured by citing the Document number and by remitting $5.00 for photoprints, or $2.25 for 35 mm. microfilm. Advance payment is required. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress.

First Order Factors

PNI Attitude Schedule. Salient variables for the first factor drawn from the PNI emerged as follows:

> PNI Factor 1—*Psychopathy.* The only way to make big money is to steal it. It's dumb to trust older people. A lot of times it's fun to be in jail. The only way to settle anything is to lick the guy. If you don't have enough to live on, it's OK to steal. I go out of my way to meet trouble rather than try to escape it. I do what I want to, whether anybody likes it or not. I would have been more successful if people had given me a fair chance. A person is better off if he doesn't trust anyone. I'm really too tough a guy to get along with most kids. (—) I would rather be at home when things go wrong. Most brothers and sisters are more trouble than they are worth. I hardly ever get excited or thrilled.

The same amoral, rebellious qualities which characterized the Psychopathy dimension of the earlier factor study (13) are evident here and leave little doubt as to factor identity. Eleven of the 13 variables which appear in the present factor loaded over .30 in the previous one, and the two remaining items (I would have been more successful if people had given me a fair chance; I hardly ever get excited or thrilled) are perfectly consistent with interpretation as psychopathy.

Composition of the second PNI factor is remarkably like that of the Neuroticism dimension isolated before (13).

> PNI Factor 2—*Neuroticism.* I don't think I'm quite as happy as others seem to be. I just don't seem to get the breaks other people do. It seems as if people are always telling me what to do, or how to do things. I often feel as though I have done something wrong or wicked. It is hard for me to act natural when I am with new people. I seem to do things that I regret more often than other people do. People often talk about me behind my back. I sometimes feel that I made the wrong choice in my occupation. (—) Most of the time I feel happy. I get nervous when I have to ask someone for a job. It seems as if I've been caught in every lie I ever told. My folks move (or used to move) from place to place a lot. (—) Most people seem to like me as soon as they meet me.

Of the 13 salient items, nine had been employed in the earlier study, and eight of these occur as markers for Neuroticism. One item previously identified with Inadequacy and four new items complete the present set of defining variables. All are sufficiently indicative of low self-esteem, dysphoric mood, tension, and guilt to justify use of the term "Neuroticism" as a label for the dimension.

The first two PNI factors strongly confirmed hypotheses regarding the structural occurrence of Psychopathy and Neuroticism, but the third PNI factor bears no relation whatsoever to the remaining hypothetical dimension, Inadequacy.

> PNI Factor 3—*Delinquent Background.* When I was going to school I played hooky quite often. In school I was sometimes sent to the principal for cutting up. My folks usually blame bad company for the trouble I get into. (—) I have never been in trouble with the law. When I was a little kid, I was always doing things my folks told me not to. I am behind at least a year in school. I have very strong likes and dislikes.

In fact, the items have no obvious pertinence to any aspect of personality. Most of them have some kind of historical-environmental significance, and the label, "Delinquent Background" implies little more than a history of delinquent behavior, possibly committed in company with other members of a delinquent group. From this point of view, the factor resembles Jenkins and Glickman's (7) "Socialized Delinquency," and from its apparent reference to early history it might be construed as a unification of the "Family Dissension" and "Scholastic Maladjustment" factors distinguished in earlier research with many of these same questionnaire items. An alternative interpretation is suggested by reflecting the factor. When loading signs are reversed, it is clear that covariation among the items could stem from a tendency to deny ever having done anything delinquent, or even naughty. A response set of this kind would be analogous to defensiveness as defined by the K-scale of the MMPI, and appropriately developed measures might be useful in improving the discriminatory effectiveness of other scales. Whether PNI factor 3 should be interpreted as historical background, momentary or dispositional response set, or as something else remains to be seen. At any rate, it does not represent Inadequacy, as we had expected, and only Psycopathy and Neuroticism have been solidly replicated as personality factors defined by this particular group of questionnaire items.

Personal Index. Loofbourow and Keys developed their instrument in an era of psychometric preoccupation with such dimensions as neuroticism and general adjustment. A great many of the items refer to neurotic symptoms, and this concentration of content, quite different from that of the PNI, probably accounts for the emergence of two factors with some kind of neurotic quality.

> PI Factor 1—*Conversion Reaction.* Do your eyes often pain you? Do you ever have the same dream over and over? Are you troubled with dreams about your play? Do you ever dream of robbers? Do you get angry so that you "see red"? Did you ever have the habit of jerking your head, neck, or shoulders? Have you often fainted away? Do you feel like jumping off when you are on a high place? Do you giggle a good deal and find you cannot help it? Do you know anybody who is trying to do you harm or hurt you? Have you the habit of biting your finger nails? Have you ever had a vision?

The first factor from the Personal Index implies a variety of conversion symptoms, along with some other hysterical tendencies, and had been correspondingly named "Conversion Reaction."

> PI Factor 2—*Hypersensitivity.* Do you ever wish that you were dead? Do you feel tired a good deal of the time? (—) Do you think people like you as much as they do other people? (—) Do you usually feel well and strong? Do you ever feel that no one loves you? Are your feelings often hurt so badly that you cry? (—) Can you do good work while people are looking at you? Do you often have a hard time making up your mind about things? (—) Can you stand the sight of blood? Do people find fault with you much? (—) Do you hear easily when spoken to?

The second dimension has a noticeable depressive component, but suggests, even more strongly, an extreme vulnerability to stress and especially to social criticism. Such frailty might easily predispose depressive feelings of a reactive kind, and sensitivity itself is conceived as the core of the factor.

> PI Factor 3—*Scholastic Maladjustment.* Are you required to take subjects that you dislike? Would you select another teacher in any of your subjects if you were permitted to? Have you often been punished unjustly? Were you ever expelled from school or nearly expelled? Are you ever punished for things you do not do? Do you often fail in subjects that you dislike? Did you ever want to run away from home? Do any of your teachers tell you that you are too noisy or talk too much? Do any of your teachers mark examinations too severely? Did you ever have a nickname you didn't like very well? Did you ever have a real fight? (—) Are you doing as much or as well in school as your parents expect you to? Do you think there should be more tryout or optional classes?

One of the factors drawn from the Personal Index contained only three pure items and will not be discussed. The items in the remaining factor, like those in the third factor of the PNI, have little to do with personality, but refer instead to external matters, principally to school and the unpleasantness of academic life. "Dislike for School" would not be an inappropriate name for the factor, but it is a trifle too specific, and the label we have chosen, "Scholastic Maladjustment," has a more fittingly general meaning.

Personal Opinion Inventory. The PNI Attitude Schedule yielded one "Neuroticism" factor. The Personal Index of Loofbourow and Keys yielded two factors related to more narrowly limited aspects of neuroticism, viz., conversion tendencies and hypersensitivity. Now the Personal Opinion Inventory, made up of items from the MMPI and from research on Naval offenders, has produced still another factor whose content can be subsumed under the general concept of neuroticism, but whose focus is more specifically upon neurotic tension and the social isolation frequently related to such dynamics.

> POI Factor 1—*Neurotic Estrangement.* At times I have a strong urge to do something harmful or shocking. No one seems to understand me. It is hard for me to just sit still and relax. Sometimes my conscience makes me do things that get me in trouble. My parents never really understood me. I feel that I have often been punished without cause. One or more members of my family is very nervous. I am so touchy on some subjects that I can't talk about them. My way of doing things is apt to be misunderstood by others. I must admit I find it very hard to work under strict rules and regulations. I have the wanderlust and am never happy unless I am roaming or traveling

> about. If I am not feeling well I am somewhat cross and grouchy. It makes me uneasy when someone does me a favor I didn't expect. I would rather go without something than ask for a favor. I feel like jumping off when I am on a high place.

Many of the items have something to do with tension or its derivatives. Most of the remaining variables refer in one way or another to misunderstanding and difficulty in open communication with other people. "Neurotic Estrangement" fails to place any particular emphasis on tension, but it summarizes the social distance implied by the variables, and suggests a neurotic basis for this state of affairs.

> POI Factor 2—*Egocentricity.* A guy who doesn't look out for himself first is a sucker. We ought to worry about our own country and let the rest of the world take care of itself. It's no use worrying my head about public affairs; I can't do anything about them anyhow. I don't blame anyone for trying to grab all he can get in this world. A person does not need to worry about other people if only he looks out for himself. Any job is all right with me, so long as it pays well. Only a fool would ever vote to increase his own taxes. There are only two kinds of women—the good and the bad. Sometimes I think I'm too nice to some people. When I am cornered I tell that portion of the truth which is not likely to hurt me. I would like to hunt lions in Africa.

The only "psychopathic" factor discussed so far, factor 1 of the PNI Attitude Schedule, contained a variety of elements which permitted unification under the general concept of psychopathy. Disregard for and distrust of authority, positive enjoyment of antisocial activity, rejection of conventional norms, isolation from close family ties, and blunting of strong positive affect were all represented. The second factor from the POI, on the other hand, most strongly emphasizes a single quality, egocentricity, but the self-centeredness is so blunt in its expression that it could hardly occur except in company with other psychopathic tendencies.

> POI Factor 3—*Social Desirability.* I would never play cards (poker) with a stranger. If I get too much change in a store, I always give it back. I am embarrassed by dirty stories. It would embarrass me to have a girl tell me a dirty story. Women should not be allowed to drink in cocktail bars. (—) During one period when I was a youngster I engaged in petty thievery. I have lived the right kind of life. (—) I would like to wear expensive clothes. The average policeman is not strict enough about the law. We ought to pay our elected officials better than we do.

The items in the third factor extracted from the POI suggest a prim morality that is too "good" to be genuine. In interpreting the third factor from the PNI Attitude Schedule, tentatively labelled "Delinquent Background," an alternative interpretation was proposed, namely that a reflection of the factor might represent defensive refusal to admit the commission of any socially unacceptable acts. A similar interpretation seems tenable for the present factor, and it has been named "Social Desirability" in the belief that a response set, and nothing deeper, has occasioned coalescence of the variables.

POI Factor 4—*Concern over Others' Opinions.* I really care whether people like me or dislike me. When people dislike me, I figure it's worth while to try to change their opinion. I often think about how I look and what impression I am making on others. I never refuse to play a game because I am not good at it. I find it hard to "drop" or "break with" a friend. It is very important to me to have enough friends and social life. It makes me angry when I hear of someone who has been wrongly prevented from voting. Before I do something, I try to consider how my friends will react to it. There is something wrong with a person who can't take orders without getting angry or resentful. It is easy for me to sympathize with someone who is always doubting and unsure about things. I regard the right to speak my mind as very important.

Some time ago, Gough (2) proposed a role-taking theory of psychopathy whose central postulate was that psychopaths are unable to see themselves from the point of view of other people. The idea led to development of a questionnaire scale for measuring role-taking capacity, and performance on the scale was related rather convincingly to role-taking ability in an experiment by Reed and Cuadra (16). The fourth factor derived from the POI appears at first to be a fairly pure representation of role-taking concern, if not ability, and it is interesting to note that the factor correlates to a modest but reliable degree (—.23) with the Psychopathy factor of the PNI Attitude Schedule. The cogency of this interpretation is weakened, however, by the fact that the average factor score for delinquents is higher than for nondelinquents and by the collective impression one gets from reading the variables that they constitute an exaggerated overstatement of concern for social approval. Again, some kind of defensive response set may be involved, but this, like the role-taking hypothesis, is difficult to assess from mere inspection of the items. At least, the statements have clear pertinence to concern over the evaluation of others, and we have chosen a properly conservative label to express this fact.

K D Proneness Scale. The relative complexity of item format in the K D Proneness Scale makes factor interpretation more difficult than for the simpler, forced-choice questionnaires discussed so far. In attempting to discern the meaning of each response tendency, one must not only consider the alternative which has been chosen or denied, but must also note relations with the other available alternatives. Thus, the first item in KD factor 1, presented below, suggests endorsement of the following kind of idea: "Whatever else policemen may do—scare you, boss you, or get something on you—they do *not* try to help you." This item, like all the KD items, is presented with the crucial alternative italicized, and the three remaining alternatives in parentheses. Signs before each item indicate whether a given italicized alternative was chosen (sign omitted) or rejected (—).

KD Factor 1—*Situational Delinquency.* (—) Most policemen try to *help you* (scare you, boss you, get something on you). (—) A boy or girl should be allowed to be his own boss when he is *21 years old* (14 years old, 16 years old, 18 years old). (—) The police *are usually very fair* (make some mistakes, favor the rich, are usually unfair). Most boys stay in school because they *are required by law to do so* (have to learn to make a living, want to go to college, like school). When not in school, I can have the most fun *around midnight* (in the mornings, in the afternoons, around noon). On my report card I usually get *some failure marks* (all honor marks, mostly good marks, fair marks). I have the most fun when I play *on my street* (in my own house, in my own yard, on the playground near my house). If a person called me a dirty name, I would *fight the person* (tell him where to get off, say and do nothing, laugh it off). When I leave school or graduate, I shall *take it easy for a while* (take any job that comes along, find a good job, go to another school or college). Going to high school is *all right for some people but not for me* (a waste of time, all right if you can take the course you want to, necessary for success). (—) If a person called me a dirty name, I would *say and do nothing* (fight the person, tell him where to get off, laugh it off). Going to school right now is doing me *more harm than good* (a great deal of good, some good, a great deal of harm). I have been *unlucky* (extremely lucky, lucky, extremely unlucky). (—) Most boys stay in school because they *want to go to college* (are required by law to do so, have to learn to make a living, like school). When I am with someone else and we want something to drink, I like to *match to see who will pay* (buy my own drink, fix it so the other person usually pays, pay for all the drinks). (—) On my report card I usually get *mostly good marks* (all honor marks, fair marks, some failure marks). Of the following sports, the one I like best to watch is a *prize fight* (baseball game, horserace, basketball game).

Few of the items in the first KD factor have clear relevance to personality as a set of internal dispositions. Most of the referents are outside the individual (policemen, school situations, social regulations, and the like). Insofar as personal characteristics are represented, they apparently constitute desires for early independence, rejection of middle-class beliefs in the value of education and hard work, and wishes just to have a good time without the annoyance of rules or constraints. Factor meaning is anything but obvious, and another psychopathic expression may be involved. The attitudes implied, however, are those commonly said to arise merely from long exposure to a delinquent subculture, with no truly psychopathic basis, and we tentatively propose the name "Situational Delinquency" as a label for the dimension.

KD Factor 2—*Irresponsibility.* In school, my friends *always get me into trouble* (almost always get me into trouble, sometimes get me into trouble, never get me into trouble). I think about what I'll do when I get out of school *not at all* (all the time, most of the time, some of the time). Going to school right now is doing me *a great deal of harm* (a great deal of good, some good, more harm than good). Whenever I get into serious trouble, other people are to blame *always* (almost always, sometimes, seldom or never). Taking part in school clubs is *very unimportant* (very important, quite important, not very important). The best teachers are the ones who are *very easy* (fairly easy, fairly hard, very hard). When I leave school or graduate, I shall *take any job that comes along* (find a good job, take it easy for awhile, go to another school or college). (—) When I leave school or graduate, I shall *find a good job* (take any job that comes along, take it easy for awhile, go to another school or college). Being successful usually means having *your name in the paper* (a big fortune, many friends, the

> respect of many people). (—) In school, my friends *sometimes get me into trouble* (always get me into trouble, almost always get me into trouble, never get me into trouble). Those who get the best jobs are usually the ones who *are the luckiest* (know the right person, are the best trained, work the hardest).

The tendencies expressed in the second KD factor include denial of personal responsibility for getting into difficulty, a lack of concern over academic and vocational success, a belief that success is tantamount to notoriety, and that it is obtained through luck, rather than effort, training, or even "pull." "Irresponsibility" summarizes these attitudes rather well, and a more central psychopathic tendency may again be represented.

> KD Factor 3—*Response Set—Central Tendency.* (—) During the past month I have worried about my family *all the time* (most of the time, some of the time, not at all). During the past month I have worried about my family *some of the time* (all of the time, most of the time, not at all). (—) I think about what I'll do when I get out of school *all the time* (most of the time, some of the time, not at all). I think about what I'll do when I get out of school *some of the time* (all the time, most of the time, not at all). The pupils who have the best attendance records are almost always *good students* (honor students, poor students, sissies). Failure is usually due to *lack of hard work* (bad habits, bad companions, lack of ability). (—) Going to school causes me to be worried and upset *most of the time* (all the time, some of the time, not at all). I usually have the best time when I do things *with two or three friends* (all by myself, with one friend, with a big gang). (—) The secret of success is *ability* (just luck, hard work, money).

It seems unlikely that the variables in KD factor 3 were drawn together by any enduring personal trait. What seems to be involved is consistency in avoiding extreme statements and in selecting moderate, "middle-of-the-scale" alternatives. This is nothing more than a response set, and the title of the factor has been phrased accordingly.

> KD Factor 4—*Attitude toward School.* Most teachers are *fair most of the time* (very fair, seldom fair, never fair). (—) The most popular boys are the ones who *almost never get into mischief* (almost always get into mischief, sometimes get into mischief, seldom get into mischief). Going to college is *all right if you can afford it* (necessary for success, all right if you have the ability, just a waste of time and money). Going to high school is *all right if you can take the course you want* (a waste of time, all right for some people but not for me, necessary for success). (—) The best teachers are the ones who are *very hard* (very easy, fairly easy, fairly hard). Cheating in school is usually done by *most of the pupils* (only a few bad pupils, none of the pupils, all of the pupils). Cheating in school is usually done by *none of the pupils* (only a few bad pupils, most of the pupils, all of the pupils).

Six of the seven items in KD factor 4 have something to do with school, but the nature of the attitude represented is difficult to stipulate. Some of the variables seem related to a vague lack of enthusiasm for educational matters, but attitudinal implications are not sufficiently specific to warrant interpretation beyond the obvious reference to academic affairs.

Intercorrelations among Factors

The complete table of correlations among first order factors can be obtained, along with other tabular materials, from the American Documentation Institute.[4] This discussion will principally concern relations between first order factors which were similarly interpreted.

Some concept implying neuroticism was invoked more often than any other in labelling dimensions. Four factors, Neuroticism from the PNI, Conversion Reaction and Hypersensitivity from the PI, and Neurotic Estrangement from the POI, were alleged to represent neurotic tendencies of one sort or another, and if these propositions are sound the factors ought to be related among themselves. In fact, the mean r for pertinent intercorrelations is .49, which is about as high as one could expect, considering the limited number of items on which factor scores were based.

Except in the case of PNI factor 1, specification of "psychopathic" factors was less clear than for "neurotic" ones. However, intercorrelations for some of the traits generally associated with the concept of psychopathy, e.g., Psychopathy itself from the PNI, Egocentricity from the POI, and Irresponsibility from the KD, are worth considering. The mean intercorrelation for those dimensions is .29. This value is lower than for "neurotic" factors, principally because our measures of Egocentricity (POI-2) and Irresponsibility (KD-2) were not closely related ($r = .17$). The correlation between Psychopathy (PNI-1) and Egocentricity (POI-2) was .34, and the r for Psychopathy and Irresponsibility (KD-2) was .35. With an N of 406, all three rs are significant beyond the .01 level, but none of them is particularly impressive.

Evaluation of the complete matrix of factor intercorrelations also requires consideration of relations among various factors within each test. These should of course be lower than for allegedly similar factors in different tests. The mean r for intratest correlations was .21, which is satisfactorily less than the mean for factors implying neuroticism, but not compellingly different from the mean r of putative "Psychopathy" factors.

The remaining dimensions were sufficiently diverse in meaning to prohibit supraordinate evaluation without additional analysis. To assess relations among them, as well as to test hypotheses regarding the coalescence of "Neuroticism" and "Psychopathy" factors, a second order factorization was conducted.

Second Order Factors

Results are given in Table 1. The salient variables are halves of the first order dimension, presented by indicating second order loadings and first order factor designations, with the letters "a" or "b" to show which half of each primary factor is involved.

Factor I is dominated by elements suggesting neuroticism. Every one of the first order constituents thought to refer to a neurotic tendency has

---

[4] *See* footnote 3.

Table 1

SECOND ORDER FACTORS

Factor I—Neurotic Delinquency			Factor II—Delinquent Background (or Defensiveness)			Factor III—Psychopathic Delinquency		
Loading	*First Order Designation*		*Loading*	*First Order Designation*		*Loading*	*First Order Designation*	
67	PNI-2a	Neuroticism	66	KD-1a	Situational Delinquency	56	KD-2a	Irresponsibility
67	POI-1b	Neurotic Estrangement	—66	POI-3a	Social Desirability	56	PNI-1b	Psychopathy
65	POI-1a	Neurotic Estrangement	63	KD-1b	Situational Delinquency	54	PNI-1a	Psychopathy
65	PI-2a	Hypersensitivity	50	PI-3a	Scholastic Maladjustment	52	KD-4b	Attitude toward School
65	PI-1a	Conversion Reaction	46	PNI-3b	Delinquent Background	51	POI-2b	Egocentricity
64	PNI-2b	Neuroticism	—45	POI-4a	Concern, Others' Opinions	49	KD-2b	Irresponsibility
62	PI-2b	Hypersensitivity	44	PNI-3a	Delinquent Background	47	POI-2a	Egocentricity
60	PI-1b	Conversion Reaction	43	PI-3b	Scholastic Maladjustment			
50	PI-3a	Scholastic Maladjustment	—40	POI-4b	Concern, Others' Opinions			
48	PI-3b	Scholastic Maladjustment						
47	PNI-3b	Delinquent Background						
44	PNI-3a	Delinquent Background						

a loading of .60 or more. Below that point, discontinuity appears in the distribution of loadings, and then a second cluster of first order elements can be seen. The latter refer to a history of delinquent activity and difficulty in school and characterize the present neuroticism factor as an alloplastic one. The neurotic tendencies, which would otherwise be indistinguishable from those among the clientele of a typical guidance clinic, are evidently expressed outwardly in this behavioral pattern; they are "acted out" against society.

Factor II is more difficult to interpret than the first one. The pattern of positive and negative loadings suggests that covariation could be determined by a consistent tendency to endorse statements implying a history of misbehavior and to deny items implying concern over society's evaluation of the behavior. The underlying attitudes are those commonly thought to occur among members of delinquent gangs, and this, in the absence of any clear expression of personality disturbance, suggested "Delinquent Background" as a label. Another interpretative possibility, however, becomes apparent when the factor is reflected, namely, that the factor could express a tendency to place oneself in a socially favorable light and deny past commission of delinquent acts. It is difficult to say which interpretation is more likely just by looking at the items, and further study is obviously needed.

Factor III unifies those first order dimensions assumed to reflect psychopathic tendencies and has been labelled accordingly. In general, the primary factors thought to refer to various facets of psychopathy appear as salient variables at the second order. In addition, half of KD factor 4, vaguely defined as "Attitude toward School," appears with a fairly high loading, suggesting that the attitude involved may represent another facet of the psychopathic viewpoint.

Together, the three second order factors define a structure remarkably similar to that developed by Jenkins and Glickman (7) from case history data. The conceptual parallels between "Neurotic Delinquency" and "Dissurbed Delinquency," "Delinquent Background" and "Socialized Delinquency," and "Psychopathic Delinquency" and "Unsocialized Aggression" are obvious, but operationally defined associations remain to be established.

In time, further research on these dimensions should proceed to extensive elaboration of antecedent and consequent relationships, but it should probably begin with the establishment of concurrent validity and more thorough examination of such purely psychometric properties of the measures as reliability and independence. Some preliminary checks with military personnel suggest that scales based on the second order factorization are not nearly as independent as one might expect from the nature of their derivation. This is true even when the cooperative elements of factors I and II are excluded, and further purification will probably be necessary.

In the present study, a premature effort was made to relate the first order questionnaire factors to three factors derived from teacher ratings and to a limited number of indicants taken from institutional files. The results were almost wholly incomprehensible, and, while this may mean that the questionnaire factors are unrelated to anything outside themselves, we are more inclined to question the adequacy of the ratings and of our own first order factor measures. In some cases, the latter were based on as few as seven items and were never derived from more than 17. Results with previous measures of Psychopathy, Neuroticism, and Inadequacy appeared rather promising (15). It seems reasonable to suppose that the present more comprehensively defined measures may also be related to crucial extratest variables and that their application may eventually lead to a substantial increase in knowledge about personality factors and the part they play in juvenile delinquency.

SUMMARY

Four questionnaires which differentiate delinquents from nondelinquents were administered to 203 training school inmates and to 203 high school boys. The items in each questionnaire were intercorrelated over the entire sample, and four separate factor analyses conducted. The resulting first order factors were individually interpreted, relations between them were examined, and a second order factorization was performed. The following results emerged.

1. Psychopathy and Neuroticism factors isolated in previous research were almost identically reproduced in the present first order analysis.
2. A factor called "Inadequacy" in the previous study did not appear in this investigation.
3. Differences in item content from questionnaire to questionnaire occasioned differences in meaning of first order factors, but many of the latter could be subsumed under second order concepts of Psychopathic and Neurotic Delinquency.
4. Most of the remaining first and second order dimensions seemed to refer to a history of delinquent behavior, and possibly to previous membership in a delinquent group, but interpretation as test-taking defensiveness was also frequently considered.

REFERENCES

1. CASTANEDA, A., MCCANDLESS, B. R., & PALERMO, D. S. The children's form of the manifest anxiety scale. *Child Develpm.*, 1956, 27, 317-326.
2. GOUGH, H. G. A sociological theory of psychopathy. *Amer. J. Sociol.*, 1948, 53, 359-366.
3. GOUGH, H. G. *Manual for the California Psychological Inventory.* Palo Alto: Consulting Psychologists Press, 1957.

4. Gough, H. G., & Peterson, D. R. The identification and measurement of predispositional factors in crime and delinquency. *J. consult. Psychol.*, 1952, 16, 207-212.

5. Hanley, C. H. An inventory of personal opinions. First Technical Report, Rehab. Research, U. S. Naval Retraining Command, Camp Elliott, San Diego, Calif., 1954.

6. Hathaway, S. R., & Monachesi, E. D. The personalities of delinquent boys. *J. crim. Law Criminol. Police Sci.*, 1957, 48, 149-163.

7. Jenkins, R. L., & Glickman, S. Patterns of personality organization among delinquents. *Nerv. Child*, 1947, 6, 329-339.

8. Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, 23, 187-200.

9. Kvaraceus, W. C. *K D Proneness Scale and Checklist*. World Book, 1950.

10. Kvaraceus, W. C. *Forecasting juvenile delinquency*. World Book, 1956.

11. Loofbourow, G. C., & Keys, N. *The Personal Index*. Minneapolis: Educational Testing Bureau, 1933.

12. Peterson, D. R., Quay, H. C., & Anderson, A. C. Extending the construct validity of a socialization scale. *J. consult. Psychol.*, 1959, 23, 182.

13. Peterson, D. R., Quay, H. C., & Cameron, G. R. Personality and background factors in juvenile delinquency as inferred from questionnaire responses. *J. consult. Psychol.*, 1959, 23, 395-399.

14. Quay, H. C., & Peterson, D. R. A brief scale for juvenile delinquency. *J. clin. Psychol.*, 1958, 14, 139-142.

15. Quay, H. C., Peterson, D. R., & Consalvi, C. The interpretation of three personality factors in juvenile delinquency. *J. consult. Psychol.*, 1960, 24, 555.

16. Reed, C. F., & Cuadra, C. A. The role-taking hypothesis in delinquency. *J. consult. Psychol.*, 1957, 21, 386-390.

17. Reiss, A. J. Social correlates of psychological types of delinquency. *Amer. sociol. Rev.*, 1952, 17, 710-718.

18. Tiffany, T. L., Peterson, D. R., & Quay, H. C. Types and traits in the study of juvenile delinquency. *J. clin. Psychol.*, 1961, 17, 19-24.

19. Chicago Area Project. *Delinquency rates in the city of Chicago for males age 10–16, during the years 1945–1951.*

PERCEPTION OF THE PARENT-CHILD RELATIONSHIP AND ITS RELATION TO CHILD ADJUSTMENT

Naomi M. Serot[*] *and* Richard C. Teevan
Smith College

One of the basic tenets of developmental psychology is the thesis that the early familial environment of the child, especially the pervading parental attitude or emotional tone of the parent-child relationship, is a fundamental factor influencing the development of personality. Clinical data offer strong support to the theory of a correlation between parent-child relationships and the nature of children's personality or relative adjustment (7, 8, 9, 14). Much research has been effected for the purpose of isolating the particular attitudes which affect the child (3, 5, 6, 15) and the qualities of personality that are the result of the specific attitudes determined (10). Often, however, this research has been inadequate or contradictory, leaving a confused picture of the relationships involved.

A review of the research offers a few explanations for the meagerness of results on this problem: Ausubel's study (1) suggests that the essential relation is that which exists between the child's perception of his familial environment and his adjustment and not, as has been thought, between expressed parental attitudes and childhood adjustment. Based on evidence in Swanson's study of delinquents (13), it was thought that, rather than use attitudes as good indicators of the nature of the parent-child relationship, it might be more useful to measure the proximity of a given child's relationship to the theoretical ideal relationship. These two possibilities suggested a third: if it is the child's perception of his familial relationships which affects his adjustment, then parental perception of the parent-child relationship may well disagree with the child's perception of the same, and, if so, the former is unlikely to be related to the child's adjustment. On the basis of these suggestions, the following hypotheses were constructed:

1. A child's perception of his parent-child relationship is correlated to his adjustment: the well adjusted child will perceive his parent-child relationship as relatively happy and close to the theoretical ideal, whereas the maladjusted child's perception of his relationship will be far from ideal.

---

[*] (Mrs. R. Michael Costello) 4884 MacArthur Boulevard, Washington 7, D. C.

Child Develpm., 1961, 32, 373-378.

2. There is little agreement between parental perception of the parent-child relationship and the child's perception of the same.

3. Parental perception of the parent-child relationship, therefore, does not correlate with the offspring's adjustment.

METHOD

Subjects

The sample of children consisted of 102 boys and girls, 9 and 10 years of age, who attended the fourth grade in public schools. The schools used were chosen because of the variability in the socioeconomic level of the parents and because of a reputed range of student adjustment. Fourth graders were used because they are young enough not to be involved in adolescent dynamics and old enough to comprehend the tests, and, also, it was thought that they might be more willing to respond truthfully than older children, more aware of social meanings, would be. The mean number of siblings was three.

The parents were those of the selected children, excluding those of children who obtained the mean adjustment score. Seventy sets (both mothers and fathers) were contacted. Most of the parents were of upper-lower and lower-middle class, with about one-sixth of lower-lower and one-sixth of middle- or upper-middle class.

Instruments

The California Test of Personality (CTP), Elementary Series AA, was used to measure the children's adjustment. A high score on the CTP indicates relatively good adjustment.

The Swanson Child-Parent Relationship Scale (CPRS) (13) was used to measure perception of the parent-child relationship. Swanson, using the scale on adolescents, found that it did not discriminate well at the ideal end of the test, but that it was a good discriminator at the other end of the scale (13). Since other relationship measures have proven to be so inadequate, the CPRS seemed to be the most usable measure in spite of this difficulty. The language and concepts of the scale were sufficiently elementary to be comprehended by fourth graders, and the form of the items was such that they could be reworded to provide a parallel measure for the parents' perception and, thus, afford a means of comparison. Consequently, the parents' scale contained the same questions, but they were reworded so that each item asked the parent for his perception of the relationship. A high score on the CPRS indicates a poor parent-child relationship as perceived by the subject.

In addition to the two tests, a short information questionnaire was used for each student. It included such items as name and address, whether or not the child lived at home with both real parents, father's occupation, and number of children in the family.

Procedure

The teachers of the three chosen classes distributed letters asking for parents' permission to use their children. The children were given no information previous to the testing time. The day the tests were administered, the children were told that the purpose of the tests was to gather information about fourth graders as a whole and that the tester was not interested in individual people. No names were requested, and the fact that names would remain anonymous was emphasized. Each test was coded so that the two tests and questionnaire of a given subject could be identified. The children were also told that their true feelings were being sought and that there were no right or wrong answers: they were to answer as they really felt. These facts were reiterated, and it was suggested that they ask questions when they had difficulty recognizing words. It was found that those children who could not recognize words could, however, understand them when they were pronounced for them. The CTP was administered first, according to the standardized instructions in the manual. Each child was allowed to finish the test at his own rate of speed. The CPRS was given next, again with no names requested and with reminders about answering according to true feelings with no fear of being identified. Thirdly, after all the tests were collected, the children filled out the information questionnaire.

The parents were sent letters explaining, nebulously, the nature of the study, requesting their participation, and giving instructions for taking the tests. Two copies of the parental form of the CPRS were included. Mothers and fathers were asked to participate individually. Names were not requested, and, again, the tests were coded so that they could be matched with children's tests. Two weeks after the letters and tests were mailed, a letter of reminder was sent out, requesting that those who had not done so return their questionnaires. Thirty-one sets of parents and five additional mothers replied adequately. Fifty-one per cent of the mothers, therefore, and 44 per cent of the fathers participated. The percentage of returns, compared to other studies (11, 12), was quite high, as much of the past research has been based on less than 20 per cent returns. The mean adjustment score of the children whose parents participated in the study was the same as the total child sample. The returns, therefore, were considered adequate.

The final parent sample consisted of 31 fathers and 36 mothers. Ten of the original children were absent the day the tests were given, one did not receive permission to participate, and three were omitted because they did not live at home with both real parents; thus, the total child sample consisted of 88 children, 44 boys and 44 girls.

RESULTS

The Pearson product-moment correlation coefficient between the children's CPRS and their scores on the personal Adjustment part of the CTP was −.77; the Social Adjustment scale of the CTP, minus the Family Relationship component (one group of items on the scale which is similar to the CPRS), correlated —.84 with the CPRS; the correlation of the CPRS with the Total Adjustment score, minus the Family Relationship component, was —.80. All three correlations are significant beyond the 1 per cent level of confidence.[1] A negative correlation means that a low score on the CTP (poor adjustment) correlates with a high score on the CPRS (perception of the parent-child relationship as far from ideal). Thus, there is significant support for the first hypothesis: the well-adjusted child perceives his parent-child relationship as relatively happy and close to the theoretical ideal, and the maladjusted child's perception of his parent-child relationship is far from ideal.

The correlation between mothers' and offsprings' CPRS scores was .17; between the fathers' and offsprings' scores, it was .19. Both correlations are insignificant and, therefore, support the second hypothesis: there is very little agreement between parental perception of the parent-child relationship and the child's perception of the same.

A correlation of .04 was found between mothers' CPRS scores and their children's scores on Total Adjustment, minus the Family Relationship component of the CTP. The correlation between the fathers' CPRS scores and the children's adjustment was .16. Both correlations are insignificant and support the third hypothesis: the parents' perception of the parent-child relationship is not related to their children's adjustment.[2]

Discussion

In an attempt to discover the limitations of the results, a chi square was determined between socioeconomic status of the family (measured by occupation) and the children's adjustment. There was no significant relation. The means for the CTP and the CPRS were the same for children in large families as they were for those in small families. Socioeconomic status and the number of children in the families were excluded as factors that could have affected the correlations. There seemed to be no other common contaminating factor which could have influenced the test results. Because of the young age of the children and the manner in which the tests were administered, it is fair to assume that the children responded truthfully. It is highly improbable that a child of 9 or 10 years of age could so alter his responses as to make his scores on both tests correlate, and it is even more improbable that a large enough part of the sample could have seen through the tests. It also seems justifiable to assume that the parents were honest, for had they not been sincere they need not have participated. Also, they often added comments to the test questions, indicating thoughtfulness and concern. Based on information given by the teachers, it is obvious that ability to answer discriminatively did not correlate with IQs.

---

[1] All correlations in this paper were calculated on two-tailed tests.

[2] The product-moment correlation coefficient for an agreement score between mothers and fathers and their children's CTP scores was .21, which is not significant. For the same sexed parent and the CTP, the coefficient was —.47; for the opposite sexed parent and CTP, it was .20. It must be remembered that two-thirds of the parents who responded were parents of boys and that boys had lower adjustment scores than did girls.

The only indicated limitation to the test results is the skewed discriminating character of the CPRS. It differentiated more adequately on this sample of children than it did on Swanson's sample, although it still was a better discriminator at the upper end of the scale. As the test was a reliable indicator of a poor parent-child relationship, it seemed that, instead of negating the validity of the correlations, the broad rather than precise differentiating characteristic of the scale merely decreased the size of the correlation. In defense of this conclusion, it is noted that, when the correlations were determined for each class separately, the greatest correlation came from the class with the most maladjusted children (independently determined).

The major implications of the results focus on the child's perception of his home life. It seems that an important developmental step has been underemphasized in theory and almost absent from research. Previous experiments have tried to relate parental attitudes or the quality of the parent-child relationship (measured by questionnaires or interviews) directly to the nature of child adjustment. They have not discovered definite one-to-one relations, for they have failed to take into account the fact that the child reacts to his perception of the situation and not directly to the situation itself. The results of this study imply that, when this fact is recognized, the expected correlations appear.

The parent-child relationship of a family, which includes parental attitudes toward childrearing, involves many subtleties which have individual, often subconscious, meanings for all concerned. These meanings influence each individual's perception of the relationship. The child's perception of the relationship is of extreme importance to him, for it is directly related to his adjustment; the actual parent-child relationship is only indirectly related to a child's adjustment. Studies of the dynamics of child psychology must, therefore, discover the course of development of children's perception of their familial environment and the factors that allow parental perception of the same environment to be so very different. Only then will it be feasible to discover more exactly the relation between child adjustment and parental attitudes.

SUMMARY

A review of the research performed within the area of parent-child relationships and child adjustment led to the formulation of three hypotheses: (a) a child's adjustment is related to his perception of his relationship with his family; (b) the child's perception of the relationship is unrelated to his parents' perception of the same; and (c) the parents' perception of the relationship is unrelated to his offspring's adjustment. The method used

to test these hypotheses was described. The statistical results obtained significantly support the hypotheses.

REFERENCES

1. AUSUBEL, D. P., BALTHAZAR, E., ROSENTHAL, I., BLACKMAN, L., SCHPOONT, S., & WELKOWITZ, J. Perceived parent attitudes as determinants of children's ego structure. *Child Develpm.*, 1954, 25, 173-183.
2. BROWN, A., MORRISON, J., & COUCH, G. The influence of affectional family relationships on character development. *J. abnorm. soc. Psychol.*, 1947, 42, 422-428.
3. BROWN, F. An experimental study of parental attitudes and their effect upon child adjustment. *Amer. J. Orthopsychiat.*, 1942, 12, 224-230.
4. BURCHINAL, L. G., HAWKES, G. R., & GARDNER, B. The relationship between parental acceptance and adjustment of children. *Child Develpm.*, 1957, 28, 65-77.
5. CHWAST, J. A study of the relationship between boys' perceptions of parental attitudes and their pre-delinquency. Unpublished doctoral dissertation, New York Univer., 1956.
6. HATTWICK, B., & STOWELL, M. The relation of parental over-attentiveness to children's work habits and social adjustments in kindergarten and the first six years of school. *J. educ. Res.*, 1936-1937, 30, 169-176.
7. JACKSON, E. B., KLATSKIN, E. H., & WILKIN, L. C. Early child development in relation to the degree of flexibility of maternal attitudes. *Psychoanal. Stud. Child*, 1952, 7, 393-428.
8. MARTIN, A. R. A study of parental attitudes and their influence upon personality development. *Education*, 1942-1943, 63, 597-608.
9. NEWELL, H. W. Psychodynamics of maternal rejection. *Amer. J. Orthopsychiat.*, 1934, 4, 387-395.
10. RADKE, M. J. *The relation of parental authority to children's behavior and attitudes.* Univer. of Minnesota Press, 1946.
11. STOGDILL, R. Attitudes of parents toward parental behavior. *J. abnorm. soc. Psychol.*, 1934-1935, 31-32, 293-297.
12. STOTT, L. H. Parental attitudes of farm, town, and city parents in relation to certain personality adjustments in their children. *J. soc. Psychol.*, 1940, 10, 325-340.
13. SWANSON, G. E. Development of an instrument for rating child-parent relationships. *Social Forces*, 1950, 18, 84-90.
14. SYMONDS, P. M. *The psychology of parent-child relationships.* Appleton-Century, 1939.
15. ZUCKER, H. J. Affectional identification and delinquency. *Arch. Psychol.*, 1943, No. 286.

ANXIETY AND VERBAL BEHAVIOR IN CHILDREN[1]

JAMES W. BARNARD, PHILIP G. ZIMBARDO, *and*
SEYMOUR B. SARASON*
Yale University

The main purpose of this study was to determine whether certain of the previous findings obtained on the Yale Test Anxiety Project could be extended to different kinds of situations. Up to the present our investigations of the effects of test anxiety (6) have been carried out in situations where the *S*s were asked to perform on a problem-solving task. Although some of the tasks were relatively unstructured, the exact nature of the desired response left unspecified, the *S* could not have helped viewing the tasks as "testing" something. Thus, for certain *S*s the experience was bound to cue off certain responses of anxiety.

It was decided, therefore, to study the effects of test anxiety on the verbalizations made by *S*s during a face-to-face interview situation in which the *S* was asked to talk about how he felt about various things ranging in generality from the things he did last summer to something that happened to him that made him very worried. To attempt to manipulate the hypothesized arousal of fears of inadequacy which are signaled as being near conscious awareness by feelings of anxiety, two interview conditions were used. It was felt that the differences predicted would be increased by indicating to the *S* (in one condition) that he would be given a test after the interview. Thus, the arousal of anxiety would occur before the interview started and would be enhanced by whatever aspects of the situation that already suggested to the *S* that he was in a test situation. Further, it was hoped (with somewhat less conviction) that it would be possible to reduce the apprehension of the *S* about the interview being a test situation by simply telling him (in a second condition) that the interviewer and he were just going to talk and that this was in no way a test.

One of the major sources of the interfering effects of anxiety on the performance of certain tasks is the compelling nature of the feeling of anxiety itself so that the individual who is experiencing anxiety is less able to attend to the external stimuli relevant to the task. He is, so to speak, forced to attend to the uncomfortable feelings of anxiety. We hypothesized that among high anxious (HA) individuals one would find a greater

---

* Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut.

[1] This study was supported from funds of an NIMH grant. The authors wish to acknowledge their gratitude to Mary Koenig, Nancy Feinstein, Lawrence Berkowitz, Kennedy Hill, and Dr. Robert Abelson for their assistance with various stages of this study.

Child Develpm., 1961, 32, 379-392.

number of references to negative feelings given in verbal productions under evaluative conditions than one would find among low anxious (LA) *Ss* under the same conditions. Further, this difference in verbal productions would not be found between the two anxiety groups under the permissive conditions. The above, of course, reflects the HA *S*'s concern over his uncomfortable feelings of anxiety. In addition to the differences in the number of intrusions of negative effect, one would expect under the evaluative condition that the responses of the HA *Ss* would be less appropriate in the direction of affect expressed to the questions asked. Thus, the LA and the HA *Ss* would differ most in the number of negative affective statements made to questions that would seem to call for positive affective statements only and in the number of positive affective statements made to questions that would seem to call for negative affective statements only.

Another difference that we expected to find was in the source of the expressed affect as indicated by the *S*. Doris (1) found that blame was assigned to others and to the self more often by HA than by LA *Ss* and that self-blame was the stronger of the two blame tendencies for the HA *Ss*. Thus, we expected that the HA *Ss* would attribute their expressed feelings to themselves to a greater extent than would the LA *Ss*.

A further effect of the HA *S*'s preoccupation with his anxiety feelings would be his lessened interest or ability to give a complete and organized response to the interview questions compared to the LA *Ss*. From this expectation we made two predictions: first, that the responses of the HA *Ss* would be less well organized, thus less comprehensible, than the responses of the LA *Ss*; second, that the HA *Ss* would include fewer details (time, place, people, etc.) in their responses than the LA *Ss*.

It was also expected that, in addition to these stylistic differences in verbal expression as a function of anxiety, measurable differences in voice quality should be related to our independent variables of anxiety level and interview atmosphere.

The main purpose of the present study was, then, to extend past findings on the influence of anxiety on performance in a testlike situation to a face-to-face interview situation where the *S*'s performance consisted of his organization of a complex verbal response given to open-ended questions.

METHOD

Children were matched on IQ, sex, and grade and randomly assigned to one of three groups which varied in anxiety level (low, medium, and high). Standardized interviews were administered to each *S* and were scored on several dimensions of stylistic performance in addition to the expression of affect. The *Ss*' expectations about the interview situation were experimentally manipulated by means of permissive or evaluative instructions. Thus, the research design is a mixed factorial arrangement: 3 (anxiety levels) × 2 (sex) × 2 (administration treatments) with four *Ss* per cell.

Subjects

From a population of over 500 third grade children tested in 11 elementary schools in Hamden, Connecticut, a sample of 48 were selected for the present study. Six months prior to the start of this study the entire sample population received the Sarason Test Anxiety Scale for Children (TASC) (6) and the Lorge-Thorndike Nonverbal IQ battery (Form A). *Ss* in the highest, lowest, and middle 15 per cent of the anxiety scale distribution were designated as high anxious (scores from 16 and over), low anxious (scores from 0 to 3), and moderately anxious (scores from 8 to 10), respectively. Eight sets of males (of three anxiety levels each) and eight sets of females (of three anxiety levels each) were finally selected to be *Ss* in the present study.

Procedure

Interview administration. The interviews, which averaged 30 minutes in duration, consisted of a series of questions with standardized probes. They were administered individually in a school room by one of two interviewers, who were, of course, unaware of the TASC score of the *Ss*. The questions varied from specific, "affect-loaded" ones like, "Tell me about a time when you had a fight or an argument with one of your good friends," to general, neutral ones like, "Tell me about the different things you did last summer." The entire interview was tape-recorded to facilitate complete and accurate reporting and scoring.

The questions were administered in the following order:

1. Tell me about the different things you did last summer.
2. What did you think about coming back to school after last summer?
3. Tell me what you think about school.
4. What are the things you like best about school?
5. What are the things that you do not like about school?
6. What do you like to play best outside of school?
7. Tell me about your teacher last year.
8. Tell me something about your best friend.
9. What do you like about two other friends (mentioned by *S*)?
10. Tell me about a time when you had a fight or an argument with one of your good friends.
11. What are the things you like to do with your father best?
12. What are the things you like to do with your mother best?
13. How would you describe your parents if you were writing about them to a classmate?
14. Tell me about something which happened to you which made you very happy.
15. Tell me about something which happened to you which made you very unhappy.
16. Tell me about a time when you were really worrying about something.
17. Tell me some things you think might make a child feel like crying, laughing, and smiling.[2]

---

[2] Because the questions about school (2-6) generally yielded stereotyped responses from all *Ss*, they were not even scored.

Since the *Ss* varied in their anxiety specific to tests and test situations, it was felt that the experimental manipulation of the *Ss*' perception of the evaluative nature of the interview situation would maximize the efficacy of the anxiety variable. Accordingly, half of the *Ss* were led to believe that the interviewer was evaluating them and that they would be given a test following the interview. These expectations were guided by the following instructions:

> You have been asked to come here today to have a test. I am testing a lot of boys and girls your age, and I will compare how well you do with how well the other children do. I will not tell your teacher or your parents how you do on the test, only you and I will know how you do. *Before* giving you the test, I would like to ask you some questions about yourself. This is not part of the test, just some questions about how you think and feel about different things. When we have finished talking, then I will ask you to play a picture and a word game, and then you will be given the test. Is that clear? You will be given a test to find out how well you do as compared to other children your age. But first, I want to ask you some questions, which are *not* part of the test.

The interview was conducted in a generally authoritative tone, with the interviewer maintaining his "distance" from *S*. Upon completion of the session, the *Ss* were told that time had run short and it would not be possible to give the test; however, since they had talked a lot already it would not be necessary anyway. The interviewer was then friendly, supportive, and encouraging to every *S*.

For the other half of the *Ss* the interview situation was quite permissive and nonevaluative in terms of instructions, the manner in which the questions were asked, and in frequent reference to *S* by his first name. The instructions stated:

> (*S*'s first name), we are talking to a lot of children your age, in your school, and in other schools. We are talking about how they think and feel about some things. Whatever we talk about will be just between the two of us. This is *not* a test. I do not want to find out if you know your lessons, we are just going to talk about how you think and feel about some things. When we have finished talking then we can play a kind of game with pictures and words. Do you understand me, (*S*'s first name)? I just want to talk to you about how you think and feel about some things.

Since spontaneous verbal expression of affective words depends, in part, on acquaintance with the words, it was necessary to assess independently of their interview responses the *Ss*' knowledge of and ability in appropriately applying affective labels within the general interview situation. Therefore, following the interview proper, each child was asked the meaning of 21 words, 12 of which were affective words (e.g., happy, excited, worry, hate) and nine of which were nonaffective words (e.g., animal, food, vehicle).

Finally, each child was asked to select which one photo from each of 21 sets of three of the Frois-Wittmann pictures (3) best expressed the affective label supplied by the interviewer (e.g., "Which man is a suspicious man?"). All of the affective words in the vocabulary test plus nine additional ones were used in this task. Each of these two tasks averaged 10 minutes in duration.

Interview Scoring

Two judges independently scored the interviews directly from the tapes. Intervals of about two months each separated the scoring of question 1 (block 1) for all *S*s and the scoring of the second block of questions (four questions) and the third block of questions (seven questions).

After pretesting with six *S*s, the following scoring categories were adopted for use in the present study:

I. *Affect*

a. *Number and type of affect statements.* Frequencies were obtained of affective statements made to questions with positive and negative affect "pull" (i.e., to questions designed to elicit positive affective responses, like "Tell me about your best friend," and those designed to elicit negative responses, like "Tell me about a time you were worrying about something"). Further, each affective statement was placed in one of five categories according to its direction (i.e., negative, positive, or predominantly mixed) and its extremity. Thus, the five categories used were extreme positive, moderate positive, predominantly mixed, moderately negative, and extreme negative.

b. *Source or origin of the affect mentioned.* While *S*s many times mentioned their own reactions to various things, they also mentioned how others felt. Thus, all affective statements could be placed into one of two categories according to the source of the affect. The source might either be the self only or some other person.

c. *Discrepancy between manifest content and affect expressed.* A 50-point scale was used ranging from "no discrepancy" to "maximum discrepancy." This was a measure of the appropriateness of the *S*'s affect to what he was talking about and was not necessarily related to the appropriateness of the affect of what he said to what the question seemed to call for.

II. *Stylistic Dimensions*

a. *Voice*

1. *Degree of animation:* a 50-point scale from "complete animation" to "complete monotone." This scale ranged from a voice which had a markedly animated quality, characteristic of one reading for a dramatic tryout, to a monotone characteristic of someone proofreading aloud.
2. *Direction of animation:* a 40-point scale from "strongly positive" (i.e., a "happy," "cheerful," "excited" voice) to "strongly negative" (i.e., a "sad," "frightened" or "depressed" voice).

3. *Forcefulness:* a 30-point scale from "forceful and persuasive" through "undistinguished" to "mousy" on which voices were rated from those with a quality of forcefulness and certainty (characteristic of someone who had something to say and wanted to be heard) to those which were weak and timid (characteristic of someone who was holding back or was speaking from inside a well-insulated room).

b. *Specificity of response:* a 30-point scale from "very specific" to "very general." A very specific response included sufficient detail to know what, where, when, how, and with whom an activity or experience was undertaken, while a very general response was a minimal answer to an interview question which was relatively unrevealing (e.g., "I went for a trip").

c. *Comprehensibility of response:* a 30-point scale from "completely comprehensible" to "mostly incomprehensible." This scale was a measure of the judges' ability to understand what the *S*s said. When the *S*s mumbled, were illogical, spoke in a very low voice (or were poorly tape-recorded), they were given "incomprehensible" scores.

d. *Concreteness vs. abstractness of descriptions of people.* *S*s' descriptions of their teachers, parents, and best friends were rated as being concrete if they dealt primarily with objective, visible physical aspects of the individuals or abstract if they dealt primarily with inferences about less superficial aspects of these people, such as personality characteristics.

RESULTS

Reliability

Two judges independently scored (direct from the tape-recordings) the dependent variables for all *S*s over the first two blocks of questions (question 1 and questions 7 to 10).[3] Separate reliability estimates were thus made for each variable and for each block. For the stylistic variables (voice forcefulness, specificity of response, and comprehensibility) Pearson product-moment correlations were computed between the independent ratings and are presented in Table 1. It is evident from inspection of the table that these variables were scored with substantial reliability, especially on the second block of questions. The increase in reliability from the first to the second block of questions can probably be attributed to a "practice" effect, with the judges improving as they established familiarity with the scoring procedures and an adequate frame of reference.

The variable of the number of affect statements made was scored with even greater reliability. On 85 per cent of the total instances, the judges were in agreement as to the presence of the expression of affect in block 1.

---

[3] The judges were, of course, unaware of the *S*s' TASC score, as was pointed up by their unsuccessful attempt to place each *S* in his correct anxiety level (LA, MA, or HA).

TABLE 1

PRODUCT-MOMENT RELIABILITY COEFFICIENTS

Variable	*Block 1*	*Block 2*
Voice Forcefulness	.51	.73
Specificity	.90	.91
Comprehensibility	.60	.80

This agreement rose to 92 per cent for block 2 and was essentially the same for instances of positive and negative affect expression.

After the reliability between judges' ratings was established on questions in blocks 1 and 2, questions in block 3 were scored by a single judge.

Checks on the Manipulation and the Experimental Conditions

In order to assess the efficacy of the manipulation of the evaluative and permissive conditions under which the interviews were conducted, three checks were used: First, the two sets of instructions were independently rated from the tapes for the effectiveness with which they were presented, and no group differences were found. Secondly, the tone of the interviewer as he conducted the interview was rated along a permissive-evaluative dimension (i.e., how well the interviewer played the role defined for him according to the condition under which he was administering the interviews). The differences between the judged tone of the interviewers toward the *S*s under the two conditions was very large ($F = 63.74$, $p < .001$) and in the appropriate direction. A further check on the quality of the interviews was the rating of the amount of direction the interviewer gave *S* over and above that demanded by the standardized interview conditions. Again, no group differences were found, so that any extraneous direction being given was randomly distributed among all groups. There was, however, a much larger difference in amount of direction given by the interviewers between the first two blocks than between blocks 2 and 3. This might reflect changes in the interviewer over time (e.g., an attempt to establish rapport initially and then a closer adherence to the interview form[4]). Thus, although no direct measure was obtained of *S*s' perception of the manipulation, it can be established that the different sets of instructions were presented with equal effectiveness, that the interviewers conducted the two types of interviews using a different tone and manner for each, and that there was no more direction given in the permissive condition than in the evaluative one.

---

[4] It is interesting to note that, although the two interviewers were of opposite sex, there were no differences between the two groups of *S*s that each interviewed on any of the dependent variables.

Since the focus of the present study was on the effects of anxiety upon the expression of verbal affect, it was thought necessary to demonstrate first that the groups did not differ in their *cognitive understanding* and *cognitive ability* in applying affect labels to affect states. Although *S*s were matched on IQs, there might still be group differences in knowledge of affective words which are not measured by a general IQ test. Thus, following the interview, a vocabulary test was given which included affect as well as nonaffect words. There were no group differences in the ability to define affective words. In order to assess the *S*s' ability in applying affective labels, the photo-judgment task was given after the vocabulary task, and again no differences were found between any of the groups. Therefore, any systematic differences between groups in the overt expression of affect during the interview are not attributable to these sources.

Affective Content of Responses

An analysis of variance showed no group differences in the total number of affective statements made to all of the questions. This total number of affective statements was then analyzed into positive and negative responses with results presented in Table 2. There are no group differences in the expression of positive affect. However, differences emerge when one looks at the expression of negative affect. Inspection of Table 2 reveals an interaction between anxiety and conditions, with the HA group expressing significantly more negative affect than the LA group under the evaluative condition ($t = 1.95$, $p < .05$, one-tailed test), while under the permissive condition the LA group expresses more negative affect than the HA group (not significant, $t = 1.24$, $p < .10$, one-tailed test). Further, the HA *S*s under the evaluative condition expressed more negative affect than did the HA *S*s under the permissive condition ($t = 2.56$, $p < .02$, one-tailed test). This interaction approaches statistical significance ($F = 2.95$, $p < .07$, $df = 2/42$) when all three anxiety groups are considered. When the divergent MA group is omitted in the analysis, the interaction between

TABLE 2

VERBAL EXPRESSION OF AFFECT

		TOTAL POSITIVE			TOTAL NEGATIVE		
		LA	*MA*	*HA*	*LA*	*MA*	*HA*
Permissive	Mean	13.9	12.9	13.5	7.4	5.4	4.8
	SD	4.6	5.7	6.4	5.5	3.7	2.4
Evaluative	Mean	14.4	16.0	15.0	5.1	4.9	8.8
	SD	4.5	6.7	5.3	3.2	2.7	4.1

anxiety and conditions is statistically significant beyond the .05 level ($F = 4.94$, $df = 1/28$).

A further refinement of this analysis was possible on the basis of whether the positive or negative affect expressed was made to questions with positive "pull" (e.g., "Tell me something which made you happy") or to those with negative "pull" (e.g., "Tell me something which made you very unhappy"). There were no differences between groups when the responses were in a direction appropriate to the question. However, significant group differences emerged when the responses were in a direction inappropriate to that called for by the affective pull of the question. Table 3 reveals a highly significant interaction between anxiety level and experimental condition for the expression of inappropriate affect ($F = 10.22$, $p < .001$, $df = 1/42$). Affect expressed is more inappropriate as anxiety increases under evaluative conditions, but more inappropriate as anxiety decreases under permissive conditions. Both trends are of similar strength.

TABLE 3

EXPRESSION OF INAPPROPRIATE AFFECT

		LA	*MA*	*HA*
Permissive	Mean	4.6	2.5	1.6
	SD	3.3	1.6	1.2
Evaluative	Mean	1.6	3.0	5.0
	SD	1.2	1.8	2.2

The number of negative affective statements made to positive "pull" questions contributed more heavily to the obtained group differences in inappropriate affect expression than did the number of positive affective statements to negative "pull" questions.

An additional analysis was done of the extremity of the affect expressed to both positive and negative "pull" questions. The differences that were found were all consistent with the above trends, though separate statistical analysis was not meaningful because of the small total number of instances of extreme negative affect. However, it is interesting to note that there were no instances of extreme negative affective statements made to positive "pull" questions (inappropriate affect) by the LA group under the evaluative conditions, while five of the eight HA *S*s under the evaluative conditions made a total of 12 such statements.

Each affective statement was scored as to its source or reference. Difference scores were then constructed for each *S* which indicated the relative frequency of attributing the source of the affect to the self or to others. An analysis of variance of these scores, contrary to our prediction, revealed no significant differences between anxiety groups. A significant difference ($F = 5.85$, $p < .05$, $df = 1/42$) was found between experimental conditions, with *Ss* under the evaluative condition indicating themselves as the source of their expressed affect relatively more often than *Ss* under the permissive condition. The *Ss* under the latter condition attributed the affect to themselves equally as often as they did to others. There were no differences related to whether or not the affect expressed was positive or negative or to whether or not the affect was given in response to positive or negative "pull" questions. No sex differences were found in the above analyses.

Stylistic Variables

The following variables were analyzed by means of an extension of Lindquist's (4) type 1 analysis of variance design which permitted partitioning of the within-group variance due to questions and blocks as well as their interactions with sex, anxiety and treatment: (a) specificity of response, (b) discrepancy between affect expressed and content of response, (c) degree and direction of voice animation, (d) voice forcefulness, and (e) comprehensibility.

No differences in specificity or affect content were obtained. While no differences were noted for degree and direction of voice animation, voice forcefulness showed a significant interaction between anxiety and sex: HA boys were less forceful than LA boys, but HA girls were judged more forceful in voice than were LA girls ($F = 4.24$, $p < .05$). No effect of experimental treatments was in evidence.

As was expected, HA *Ss* were rated more incomprehensible than LA *Ss* ($F = 3.75$, $p < .05$), and the linear trend observed among the three anxiety groups was highly significant ($F = 7.39$, $p < .01$). This difference was slightly (not significantly) greater under evaluative than permissive conditions.

A final stylistic dimension of performance suggested by the research of Sigel[5] on conceptual modes of verbal expression is the use of concrete versus abstract descriptions of people important to *S* (e.g., parents, teacher, best friend). Table 4 presents a comparison of the number of concrete and the number of abstract descriptions given by *Ss* during the interview. A "concrete" description is one in terms of overt, objective, visible, physical characteristics (e.g., height, weight, hair, facial appearance). "Abstract" refers to descriptions which require *S* to make an inference about some attribute of the person, usually a personality characteristic. These descriptions are thus more subjective, less superficial, and generally deal with traits which are more important for social interaction than those mentioned in concrete descriptions.

Difference scores between the number of abstract and the number of concrete descriptions used by each *S* were computed. An analysis of variance of these difference scores, omitting the MA *Ss*, yielded significant differences between anxiety groups and near significant differences between conditions. The LA *Ss* tended to be more abstract than concrete in their descriptions; the HA children tended to be relatively more concrete ($F = 5.92$, $p < .025$, $df = 1/28$). Similarly, under permissive conditions relatively more abstract descriptions were given than under evaluative conditions ($F = 3.24$, $p < .10$, $df = 1/28$). Table 4 presents the scores representing the *difference* between use of abstract and concrete descriptions for the three anxiety groups under each experimental condition. Since approximately the same total number of descriptions were given by *Ss* in each group, it can be concluded from this table that LA, permissive *Ss* describe other people in terms of abstract characteristics, while HA, evaluative *Ss* use relatively more concrete characteristics to describe other people. Thus, both anxiety and the experimental manipulation determine whether children will tend to prefer abstract or concrete descriptions of people who are important to them. With the inclusion of the MA group the analysis yielded nonsignificant results due primarily to the divergent scores of the MA *Ss* under the permissive condition. Again, no sex differences were found.

---

[5] Personal communication, 1959.

TABLE 4

ABSTRACT VERSUS CONCRETE MODES OF DESCRIPTIONS*

		LA	*MA*	*HA*
Permissive	Mean	17.6	12.8	14.4
	SD	3.4	6.0	4.6
Evaluative	Mean	15.1	14.8	12.6
	SD	2.4	3.5	2.5

* Scores represent the difference between abstract and concrete descriptions; thus, larger numbers indicate more abstract descriptions.

DISCUSSION

As has been elaborated in a previous publication (6), we have accepted the conception of anxiety as a signal to an individual that the strength of certain thoughts or feelings that he may be experiencing have increased to the point that the behavior they might elicit is unacceptable and dangerous. This signal has at least two major properties (2): it is a phenomenon of which the child is acutely conscious, and it is of a decidedly uncomfortable nature. The situations that may cue off anxiety are naturally varied and differ for each individual. Our *Ss* were selected on the basis of how much anxiety they admitted experiencing in the school situation. One of the two experimental conditions used in this study represented an attempt to reinforce the effects of a very important aspect of this situation, the process of being evaluated. Thus, it was felt that, if a child admitted to a concern over tests and testing situations, any indication that he was in fact faced with the prospect of being thrust into such a test situation would increase his anxiety. This would have certain effects on his behavior. Specifically, the major expected effects would be intrusions of negative feelings into his verbal productions and also a lessened ability to bring together thoughts into comprehensible verbal response. The first part of the expectation was borne out in that it was found that the HA *S*s made more references to negative affect under the evaluative condition than did the HA *S*s under the permissive condition and also more than the LA *S*s made under the evaluative condition. It was further found that this difference was accentuated when the questions put to the child were judged to be positive "pull" questions. When negative affect was appropriate (i.e., as responses to negative "pull" questions), very little difference between groups was found in the number of such statements made. This might suggest that the affect expressed by HA *S*s under the evaluative condition tends, due to the pressure of inner feelings, to be less appropriate to the specific external situation. The specific measure of affect-content discrepancy used in the study failed to show differences that would support the above. Actually, this measure seems to have confounded two kinds of discrepancy: first, the discrepancy between the affect expressed in the verbal response and what would normally be called for by the question itself; and second, the discrepancy between the affect expressed and the actual content of the verbal response, disregarding what the question would seem to call for. Such a confounding could conceivably lead to the above lack of results.

It can be noted from Table 2 that, not only do the HA *S*s under evaluative conditions make a large number of negative affective statements, as was initially expected, but also the LA *S*s under permissive conditions make a large number of such statements. This finding was surprising, and we have no ready explanations. One conceivable explanation of this unexpected result might lie in the possibility that the tendency to deny anxiety feelings was unevenly distributed among *S*s in the various groups. It has long been our opinion that the measurement of the defensive tendency is the bottleneck in the development of more effective personality screening procedures. Ruebush (5), using a measure of defensiveness, found, that among *S*s who received low TASC scores, those who also received high defensiveness scores performed more like HA *S*s and less like LA *S*s who received low defensiveness scores. This further breakdown of levels of anxiety into categories of defensiveness will be followed up in a subsequent study.

Additional differences were expected in the specificity of the *S*s' responses. Only a trend barely suggestive of an interaction between anxiety and sex was found: the boys scored in the expected direction, that is, the LA boys were more specific in their responses than were the HA boys. This disappointing result might be due to the fact that the measure itself was too broad and perhaps brought together opposite trends by giving the same weight in the final score to each of the various details looked for.

Analyses of the three voice categories showed that the pattern of results for all three were similar but with only the differences between groups in the voice forcefulness category reaching significance. As with the measure of specificity, an interaction between anxiety and sex was found with the trend among the boys being in the expected direction (i.e., LA boys being judged more forceful than the HA boys) while the trend among the girls was in the reverse direction. This sex difference parallels other findings where the HA girls actually seem to perform, at least in some ways, in a manner more similar to the LA boys than to the HA boys. Very probably there are differences in inferences to be drawn from a girl's admitting experiencing anxiety in the school situation than a boy's admitting the same thing. For girls such feelings are more acceptable socially. In fact, a girl who denies having any feelings of anxiety might almost be thought to be stepping out of her expected role. In any event, the systematic investigation of the measurement and subsequent effect of anxiety in behavior as it differentially influences the two sexes is another area deserving more attention.

Over-all, *S*s in the evaluative condition gave relatively more concrete responses than *S*s in the permissive condition. Also, HA *S*s in the evaluative condition gave relatively more concrete responses than did the LA *S*s in the same condition. When this last finding is viewed in the context of our other results (e.g., comprehensibility, appropriateness of affect, etc.), it suggests that the marked concreteness of the HA *S*s reflects in some way the effects of interfering factors in performance under stress. It is our opinion, based on previous studies, that the concreteness of the HA *S*s is a reflection of their prepotent tendencies to respond cautiously.

The inclusion of a moderate anxiety (MA) group was intended to supply information as to whether TASC score was linearly related to our various dependent variables. Unfortunately, at least for the sake of simplicity, more questions seem to have been raised than have been answered. On some of the measures, notably comprehension, it was demonstrated that there exists a linear trend over all three anxiety groups. However, with other measures such as abstract vs. concrete descriptions of people and number of affective statements made, it was found that the MA group's score fell below or above the LA and HA group scores. At this point, such disparate results are difficult to explain. However, it may very well be that the MA group is made up in general of individuals whose performance on certain tasks is better than the performance of the other two anxiety groups due to their specific mode of handling feelings of anxiety. On the one hand, because they have moderate scores on the TASC, they cannot be thought of as generally denying their concerns of the school situation; but yet their TASC scores are low enough to indicate that probably they are able to handle such concerns to the extent that they would not interfere with their performance on school tests. Again, the necessity is pointed up of employing a measure of defensiveness to further isolate the relevant factors at work both in determining an individual's TASC score itself and in determining the effects on behavior of the anxiety that is reflected in that score.

SUMMARY

In order to study the relation between anxiety and certain aspects of verbal behavior (especially expression of affect) standardized interviews were administered to 48 third grade children of both sexes, matched for IQ, but varying in their level of test anxiety. The *S*s' perception of the interview situation was manipulated by evaluative or permissive instructions. The evaluative treatment was designed to cue off anxiety, and the permissive treatment was designed to alleviate anxiety.

The major findings were that HA children in an evaluative interview situation express more negative affect than LA children in the same situation, and that this difference does not exist under a permissive interview situation. This difference is accentuated when the expression of negative affect is inappropriate to the stimulus question. Further, this difference seems to become greater as the negative affect becomes more extreme. In describing other people who are important to him, the anxious child tends to be relatively more concrete in his descriptions than the nonanxious child, and this difference is enhanced by the experimental treatment. The responses of the HA *S*s were more incomprehensible than those of the LA *S*s under both experimental conditions. For the variable of voice forcefulness there was a significant interaction between sex and anxiety, with the HA boys being least forceful, while the HA girls were the most forceful.

REFERENCES

1. DORIS, J. Test anxiety and blame assignment in grade school children. *J. abnorm. soc. Psychol.*, 1959, 58, 181-190.
2. FREUD, S. *The problem of anxiety.* Norton, 1936.
3. HULIN, W. S., & KATZ, D. The Frois-Wittman pictures of facial expression. *J. exp. Psychol.*, 1935, 18, 482-498.
4. LINDQUIST, E. F. *Design and analysis of experiments in psychology.* Houghton Mifflin, 1953.
5. RUEBUSH, B. K. Children's behavior as a function of anxiety and defensiveness. Unpublished doctoral dissertation, Yale Univer., 1960.
6. SARASON, S. B., DAVIDSON, K. S., LIGHTHALL, F. F., WAITE, R. R., & RUEBUSH, B. K. *Anxiety in elementary school children.* Wiley, 1960.

RELATION OF OBJECT CURIOSITY TO PSYCHOLOGICAL ADJUSTMENT IN CHILDREN[1]

PAUL McREYNOLDS* *and* MARY ACKER
Veterans Administration Hospital, Palo Alto

and CARYL PIETILA[2]
Portola Valley School, California

During recent years there has been considerable research interest in the related topics of exploratory behavior, curiosity, and the tendency to seek and maintain novelty in perception. Most of the experiments in this general area have dealt with subhuman species: thus, Harlow and his collaborators (7, 8) have emphasized the importance of "curiosity-investigatory" motives in monkeys; and Montgomery and his associates (17, 18, 19) have demonstrated the motive power of what they term the "exploratory drive" in rats. There have been only a few studies concerned with curiosity in humans: Berlyne (1, 2, 3) investigated the stimulus factors related to curiosity in adults; Kagan, Sontag, Baker, and Nelson (10) found IQ changes in children to be related to curiosity themes on the TAT; and one of us (14) presented tentative evidence that anxiety and curiosity are negatively related in psychiatric patients. For brief reviews of research in the whole area of exploratory behavior the reader is referred to Berlyne (4) and Butler (5).

The manifest fact of curiosity in humans has been taken into account by many theorists, including James (9), McDougall (12), Troland (20), Linton (11), Maslow (16), and the psychoanalysts (6). One of the present authors has proposed a conceptualization of human motivation which to a considerable degree is built around the tendency of individuals to seek novelty in their perception. This point of view, which is more fully described

---

* Veterans Administration Hospital, Palo Alto, California.

[1] This study was made possible through the assistance and cooperation of Mr. Anthony L. Rose, Superintendent of Portola Valley School District.

[2] Shortly after the completion of the collection of data for this study, Miss Caryl Pietila was killed in an automobile accident. She was an understanding teacher, a gifted research worker, and an outstanding person.

Child Develpm., 1961, 32, 393-400.

elsewhere (13, 14, 15), assumes that persons have inherent tendencies to seek novel percepts. It also assumes that persons tend to assimilate incoming percepts into their apperceptive masses. The concept of assimilation refers to the incorporation and integration of incoming percepts into harmonious and concordant relationships with existent perceptual systems. The process of assimilating the continuing input of percepts which one receives in the course of living would be expected to proceed smoothly as long as there were no major incongruencies between incoming percepts and the schemata according to which they would be assimilated, but in the presence of outstanding incongruencies assimilation would be difficult if not impossible. Thus, material about which one is ambivalent or which represents problem areas for him would be difficult to assimilate. It is assumed that unassimilated percepts tend to accumulate and that anxiety is a positive function of the quantity of unassimilated perceptual material. The theory proposes that the tendency to obtain novel percepts—in other terms, curiosity or exploratory behavior—is attenuated as the magnitude of unassimilated percepts becomes greater.

A test of this hypothesis with children as subjects seems particularly appropriate because of the generally accepted view—which perhaps has the status of an item of folk wisdom—that curiosity is particularly prominent in childhood. In order to test the hypothesis directly it would be necessary to measure both curiosity and magnitude of unassimilated percepts. Both of these present considerable difficulties. With regard to curiosity it is not known whether or to what degree curiosity is a general trait, and it is conceivable that a child might be curious concerning one area of the environment and incurious concerning another. Or he might express curiosity for objects but not for abstract ideas, and so on. The technique which we used in the present study to measure curiosity concerned exploratory behavior with objects, and, while it is our presumption that this measure is at least somewhat indicative of curiosity generally, it seems best, in the interests of scientific accuracy, to refer to the subject of our measure as *object curiosity.*

Rather than attempting to measure directly the quantity of unassimilated percepts—or anxiety, which this quantity is held to underlie—we obtained estimates of several different aspects of psychological adjustment. This procedure was based on the assumption that in general a child who is maladjusted has a greater amount of unassimilated material than a child who is well adjusted. Such an assumption seems quite reasonable: the maladjusted child is likely to be characterized by emotional ambivalences which make difficult the assimilation of percepts in important areas of experience; that is, he is likely to have undergone experiences which he has not been able to work through or to integrate into his personality. It is true, of course, that on occasion well-adjusted children would be expected to have considerable unassimilated perceptual material, but in general the present assumption seems justified.

The specific hypothesis tested in the present study was that object curiosity in children is negatively related to psychological maladjustment, or, to word it differently, that curiosity is positively related to adequacy of adjustment. This prediction, while relevant to the conceptualization from which it was derived, clearly is not crucial to that particular theory and has implications beyond that theory. For example, the question as to whether curiosity is inhibited by psychological maladjustment is of immediate significance to educators.

METHOD

Subjects

Thirty children, 14 girls and 16 boys, of mean age 11.5 served as subjects. These students comprised the sixth grade class at Portola Valley Elementary School, located in an upper middle class community near Palo Alto. The achievement level of this class as a whole may be estimated from results of a May 1959 administration of the California Achievement Tests which placed approximately 88 per cent of the class above sixth grade level and 40 per cent above eighth grade.

Procedure

Data were gathered from two main sources: ratings made by the sixth grade teacher on 10 behavioral variables and curiosity scores obtained from observing each *S* in a special test situation. The examiner (MA) who administered the curiosity test was introduced to the group by their teacher (CP) on a visit to the school prior to the testing period. At that time, the class was told that the purpose of the study was to see how well they could identify objects by manipulating them without looking at them, but curiosity *per se* was not mentioned.

Measure of object curiosity. The technique used for measuring object curiosity was a modification of that used in previous studies (14). Essentially, it involved a systematic observation of *S*'s spontaneous exploratory behavior when presented with a number of interesting objects.

The experimental material consisted of the following: (a) the Object Box, a green wooden box 18 in. by 10 in. by 10½ in. in size, covered in front with heavy green curtains which parted at the center and open at the back; (b) the Free Play Box, a green wooden box 17 in. by 17 in. by 3¼ in. in size; it was open at the top and consisted, in egg-crate fashion, of 16 cubicles, each 4 in. by 4 in. by 3 in. in size; and (c) thirty-five small objects, such as a miniature camera, a small stapler, a harp, a toy motorcycle, and so on.

Individual sessions were scheduled 30 minutes apart. The testing sequence began as each *S*, when his turn came, was escorted from his classroom by *E* to a small separate building, sometimes used as a library and familiar to the students. In the testing room, *S* was seated at a table facing the Object Box, with *E* sitting around the corner of the table to *S*'s right.

The experimental session consisted of two phases. First, there was an identification task during which *E* placed 12 objects, one at a time, inside the Object Box. She did this by inserting the object through the open back of the box so that the child could not see the objects. *S* reached his hand into the box, through the opening in the curtain, and attempted to identify each object solely on the basis of cues received from feeling it and manipulating it. After each guess, *S* was free to take the object out of the box and look at it or play with it until the next object was put in the box for his identification. E unobtrusively observed the extent of *S*'s exploratory behavior with each object, according to a detailed, predetermined schedule on which possible manipulations of each object were listed. For example, as an *S* observed the miniature flashlight, one point was scored for each of the following manipulations: move top to different angle, unscrew top, unscrew bottom, remove bulb, clip onto something, install battery, turn on, read writing on side. In addition to scoring various kinds of manipulations of an object, one point was given for each meaningful question about an object, and one point was given for each comment that implied curiosity (e.g., "These flashlights are manufactured in Germany.") After the 12 objects had been presented, *S* was told that the "test" was over and that, while *E* prepared scoring sheets and objects for the next child, *S* would be free to amuse himself for a few minutes. At this time, *S* was shown to another table on which there was a box (Free Play Box) containing 23 additional objects which he might examine while waiting if he cared to do so. The period of free play that followed was timed by *E* for 10 minutes, during which time she unobtrusively watched *S*'s manipulations of the objects and recorded score points according to predetermined criteria, as illustrated by the miniature flashlight example above.

The *object curiosity score* consisted of the sum of all points earned by *S* in exploring the 35 objects. An estimate of the reliability of this score was obtained by the split-half method. The objects were divided into two groups, matched as closely as possible in terms of (a) the number of possible points which could be earned for each of the objects (e.g., the real miniature camera had many possible responses whereas the rubber screw driver had few), (b) whether the objects were used in the identification or the free play phase, and (c) the extent to which the objects appeared to be of primary interest to boys or to girls. The correlation between the two halves, as corrected by the Spearman-Brown formula, was .87. This coefficient is of the order of that (.96) reported previously (14) for a similar procedure. It does not directly concern the reliability of scoring, but this can be presumed to be quite high due to the detailed nature of the scoring criteria. In the one of the previous studies (14) in which two judges were used, the correlation between scores assigned by the two judges was .99.

Ratings of behavior. The estimates of psychological adjustment and maladjustment were based upon ratings made by the classroom teacher. Because of the complex nature of psychological adjustment, six different behavioral rating scales were utilized, as follows: nervous behavior, worry over achievement, classroom adjustment, adjustment to teacher, adjustment to peers, and over-all psychological health. In addition, and for general interest rather than for specific pertinence to the hypothesis under study, ratings were obtained on scholastic motivation, activity level, creativity and originality, and curiosity, with the prediction in each case being a positive relation with object curiosity. While the rater, as a co-author, was of course familiar with the general design of the study, she had not seen and had no direct knowledge of the method of administering or scoring the object curiosity test. The ratings were made in the latter part of the school year so that the teacher was quite familiar with each student. Ratings were made on six-point scales, with the 30 *S*s distributed as much as possible about the six points so as to approximate a normal distribution (1:5:10:10:5:1). Detailed definitions of the behavioral variables were prepared for the rater's guidance.[3]

The fact that only one rater was used is a shortcoming in the present study, but the fact is that no persons other than the teacher were sufficiently familiar with the *S*s to evaluate them. Some indication of the reliability of the ratings can be obtained by comparing the ratings assigned on closely similar variables: thus, "creativity and originality" appears to be logically similar to "curiosity," and these two ratings were correlated .76. Similarly, "maladjustment in classroom" and "maladjustment to teacher" were correlated .83.

RESULTS AND DISCUSSION

Table 1 presents the correlations between object curiosity and the behavioral variables. It is obvious that our primary hypothesis of a positive relation between object curiosity and psychological adjustment and of a negative relation between curiosity and maladjustment is supported. While of modest magnitude, all of the correlations are in the predicted direction, and most of them are statistically significant. The findings thus give indirect support to the theoretical interpretation of a negative relation between amount of unassimilated material and strength of the tendency to obtain novel percepts.

From a more immediately practical point of view, the results raise the possibility that those aspects of classroom learning which depend upon curiosity are hindered by the anxieties of the student.[4] This conclusion appears to have significant implications for educational practice.

---

[3] Duplicated copies of the rating scales, as well as lists of the objects presented for exploration and the detailed instructions to the *S*s, can be obtained from the authors upon request.

[4] This tentative conclusion is different from the common view that maladjusted people are often very creative, indeed that maladjustment somehow contributes to creativity. So far as we are aware, however, this opinion is unencumbered by systematic studies; it is a reasonable hypothesis that, whereas severe maladjustment may sometimes increase motivation for creative productivity, it does not directly facilitate originality.

TABLE I

PRODUCT-MOMENT CORRELATIONS BETWEEN PSYCHOLOGICAL ADJUSTMENT AND OBJECT CURIOSITY ($N = 30$)

Behavior Variables	*Object Curiosity*
Psychological Adjustment:	
Over-all Psychological Health	.45*
Psychological Maladjustment:	
Nervous Behavior	—.42*
Worry over Achievement	—.27
Maladjustment in Classroom	—.50*
Maladjustment to Teacher	—.42*
Maladjustment to Peers	—.39**
Other Variables:	
Scholastic Motivation	.43*
Activity Level	.03
Creativity & Originality	.20
Curiosity	.37**

NOTE.—The data for Maladjustment to Peers and Over-all Psychological Adjustment are based on 29 cases.

* Significant at .01 level, one-tailed test.

** Significant at .05 level, one-tailed test.

Curiosity is not necessarily a significant factor in all aspects of classroom learning, however. For the present study we had available California Achievement Test scores obtained about two weeks before the curiosity data. When these scores were correlated with object curiosity data, a product-moment coefficient of .14 was obtained. Presumably aptitude for and skill in rote learning and memory are more important than object curiosity in those aspects of learning measured by the California Achivement Tests. It is also possible that curiosity, but of a sort other than object curiosity, is significant for scholastic achievement as measured by this test.

The present study, while productive of positive findings, should be considered as preliminary rather than as definitive. It was limited by the fact that the measures of adjustment were based completely upon the judgments of one rater and by the fact that the kind of curiosity sampled was rather restricted. The study does indicate, however, that curiosity can be successfully studied in children, and it is hoped that more investigations of this variable will be forthcoming.

SUMMARY

Thirty children in a sixth-grade class were administered a test of object curiosity. The classroom teacher rated each student on a number of variables,

most of them concerned with psychological adjustment. The hypothesis was that object curiosity would be related negatively to maladjustment and positively to adequacy of adjustment. The results supported the hypothesis.

References

1. Berlyne, D. A theory of human curiosity. *Brit. J. Psychol.*, 1954, 45, 180-191.
2. Berlyne, D. An experimental study of human curiosity. *Brit. J. Psychol.*, 1954, 45, 256-265.
3. Berlyne, D. Conflict and information-theory variables as determinants of human perceptual curiosity. *J. exp. Psychol.*, 1957, 53, 399-404.
4. Berlyne, D. The present status of research on exploratory and related behavior. *J. indiv. Psychol.*, 1958, 14, 121-127.
5. Butler, R. Exploratory and related behavior: a new trend in animal research. *J. indiv. Psychol.*, 1958, 14, 111-121.
6. Fenichel, O. *The psychoanalytic theory of neurosis.* Norton, 1945.
7. Harlow, H. Mice, monkeys, men and motives. *Psychol. Rev.*, 1953, 60, 23-32.
8. Harlow, H., Blazek, H., & McClearn, G. Manipulatory motivation in the infant rhesus monkey. *J. comp. physiol. Psychol.*, 1956, 29, 444-448.
9. James, W. *Principles of psychology.* Holt, 1890.
10. Kagan, J., Sontag, L., Baker, C., & Nelson, V. Personality and IQ change. *Amer. Psychologist,* 1957, 12, 364. (Abstract)
11. Linton, R. *The cultural background of experience.* Appleton-Century, 1945.
12. McDougall, W. *An introduction to social psychology.* Methuen, 1907.
13. McReynolds, P. A restricted conceptualization of human anxiety and motivation. *Psychol. Rep. Monogr. Suppl.*, No. 6, 1956.
14. McReynolds, P. Exploratory behavior as related to anxiety in psychiatric patients. *Psychol. Rep.*, 1958, 4, 321-322.
15. McReynolds, P. Anxiety, perception, and schizophrenia. In D. Jackson (Ed.), *The etiology of schizophrenia.* Basic Books, 1960. Pp. 248-292.
16. Maslow, A. *Motivation and personality.* Harper, 1954.
17. Montgomery, K. The relation between fear induced by novel stimulation and exploratory behavior. *J. comp. physiol. Psychol.*, 1955, 48, 254-260.
18. Montgomery, K., & Monkman, J. The relation between fear and exploratory behavior. *J. comp. physiol. Psychol.*, 1955, 48, 132-136.

19. MONTGOMERY, K., & ZIMBARDO, P. The effect of sensory and behavioral deprivation upon exploratory behavior in the rat. *Amer. Psychologist,* 1957, 12, 461. (Abstract)

20. TROLAND, L. *The fundamentals of human motivation.* Van Nostrand, 1928.

RELATION BETWEEN ANXIETY AND TWO MEASURES OF SPEED IN A REACTION TIME TASK[1]

David S. Palermo*
University of Minnesota

A study by Wenar (16) reported a positive relation between anxiety, as defined by the Taylor manifest anxiety scale (MAS) (14), and response speed in a simple reaction time task. This finding was interpreted as indicating that anxiey has motivational properties and as supporting Hullian theory (6) which states that, if the habit factor is held constant, increases in drive level will lead to increases in response strength.

A number of attempts have been made to replicate Wenar's results (1, 4, 5, 7, 8, 12). With the exception of a recent study by Rossi (12), none of these studies report results confirming those of Wenar.

All of the studies have not measured the same responses as reaction time. For example, the Farber and Spence report includes two experiments; in one they measured the amount of time to move the finger from one position to another after the stimulus was presented, and in the other they measured the amount of time from the onset of the stimulus to removing the finger from a button. Some studies have measured the amount of time from onset of stimulus to starting an action (1, 4), another has measured the amount of time to make a movement from one position to another after the stimulus has been presented (4), while most of the studies have used a combination of both as a single measure (5, 7, 8, 12, 16).

---

* Institute of Child Development, University of Minnesota, Minneapolis 14.

[1] This study was supported by a grant from the Graduate School of Southern Illinois University and was conducted while the author was a member of the Department of Psychology there. The author wishes to express appreciation to Mr. Ronald Brown for his assistance in collecting and analyzing the data and to all the principals and teachers of the Murphysboro Public Schools for their cooperation in making this study possible. This paper was read, in part, at the Amer. Psychol. Ass. meeting, Washington, D.C., 1958.

Child Develpm., 1961, 32, 401-408.

The two separate measures of reaction time seem analogous to the two separate measures of response evocation or starting time and running time taken for rats in a runway. In the case of the runway situation, Spence (13) has pointed out that starting time is much more likely to be affected by competing responses and, therefore, is not as reliable a measure as running time. It would seem reasonable to assume that the same factors might operate in the reaction time task.

The present study is an attempt to determine the relation between anxiety and these two measures of reaction time and, in addition, the relation between the two measures, reaction time (RT) and movement time (MT).

METHOD

Subjects

The *S*s were 113 sixth grade children, 57 boys and 56 girls, in a public school. All of the children had taken the children's form of the Taylor manifest anxiety scale (CMAS) (2). There were 13 boys and 13 girls with scores of 10 or below (low anxiety), 31 boys and 29 girls with scores between 11 and 23 inclusive (medium anxiety), and 13 boys and 14 girls with scores of 24 or higher (high anxiety).

Apparatus

A modified reaction time apparatus made by the Lafayette Instrument Co. was used. It consists of a platform 18 in. long and 1¾ in. wide divided in the middle by a perpendicular board 1 in. thick, 10 in. high and 11¾ in. wide. On *S*'s side of the apparatus was located a microswitch mounted 1 in. above the level of the platform. In the center of the platform, 2 in. above the platform level and 4½ in. from the front edge of the platform, was located a round ¾ in. diameter button attached by a lever to another smaller button originally used as the response button on this apparatus. On the perpendicular board facing *S* were two 1½ in. square pieces of glass, one on each side, 7½ in. above the platform and 2¼ in. from the side of the board. The glass on *S*'s right was white, and the one on the left was red. Each could be illuminated from behind by a 6.2 volt G. E. flashlight. Only the red light was used in this experiment. On *E*'s side of the perpendicular board was located a three-way switch which was set so that the red light would be activated when *E* pressed the starting button. A standard electric timer was connected into the circuit such that it was started by *E* at the same instant as the onset of the stimulus and stopped at the same instant that *S* lifted his finger off the microswitch. A second standard electric timer in the circuit started at the same time the first one stopped and stopped when *S* pressed the response button turning off the stimulus. Thus, there was a measure of the time to start (RT) and a measure of time to move from one button to the other (MT).

Procedure

All *Ss* were told that the purpose of the experiment was to see how fast they could move their finger from one button to the other. They were told to place the index finger of their preferred hand on the microswitch when *E* said "Ready," and, when the stimulus went on, they were to turn it off by pressing the button as quickly as they could. The *E* gave the ready signal approximately 2 sec. prior to the presentation of the stimulus. All *Ss* were given three preliminary practice trials to insure their understanding of the instructions. There followed immediately 40 trials for which *E* recorded both RT and MT. The *S* was informed of his time scores following each trial.

RESULTS

A reciprocal transformation was performed on all RT and MT scores in order to help eliminate the skewness of the distribution of scores and to make the results comparable to previous studies. Thus, all the analyses are in terms of speed of response.

Table 1 presents separately the means and standard deviations of the RT speed scores for both sexes and the three levels of anxiety. An analysis of variance of these data over 10 blocks of four trials indicated that boys responded significantly faster than girls and that there was a significant increase in speed over trials but there were no significant effects of anxiety nor interaction effects.[2]

TABLE 1

SPEED OF STARTING SCORES FOR BOYS AND GIRLS AT THE THREE LEVELS OF ANXIETY

	LEVELS OF ANXIETY					
	LOW		MEDIUM		HIGH	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Boys	4.3	.7	4.3	.7	4.4	.7
Girls	4.2	.7	4.0	.8	3.7	.8

Table 2 presents separately the means and standard deviations of the MT speed scores for both sexes and the three levels of anxiety. Analysis of variance of these data over 10 blocks of four trials yielded the same results as in the case of the RT scores. There were significant differences

---

[2] To allow for proportional groups in the analysis of variance, two boys from the middle anxious group and one girl from the high anxious group were randomly selected and omitted from the analysis.

TABLE 2

SPEED OF MOVEMENT SCORES FOR BOYS AND GIRLS AT THE THREE LEVELS OF ANXIETY

	LEVELS OF ANXIETY					
	Low		Medium		High	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Boys	6.1	1.2	6.0	1.4	5.9	1.8
Girls	5.5	1.1	5.5	1.7	4.7	.7

attributable to sex and trials but no significant anxiety effects nor any interaction effects.

Table 3 presents the Pearson product-moment correlations between the mean RT and MT speed scores for both sexes and the three levels of anxiety. Within the groups only the correlation for the female high anxiety group reaches a statistically significant level of confidence ($p < .05$). The total group correlation of .33 is, however, highly significant ($p < .01$).

To provide an additional check on the relation between the RT and MT measures, Pearson product-moment correlations were run between the 40 RT and MT speed scores obtained for each *S*. The distribution of the 113 correlations ranged from —.68 to .60 with a mean of —.03. Table 4 indicates the mean correlations, based on Fisher's *z* transformation, obtained for the various groups in this study. It is apparent from the table that there is no significant relation between the RT and MT measures for any group or for the total group. While the distribution of these correlations appears relatively symmetrical, it tends to be platykurtic in form. Thirty-eight of

TABLE 3

CORRELATIONS BETWEEN THE MEAN RT AND MT SCORES FOR EACH SEX AT THE THREE LEVELS OF ANXIETY

	LEVELS OF ANXIETY					
	Low		Medium		High	
	N	*r*	*N*	*r*	*N*	*r*
Boys	13	.54	31	.22	13	.27
Girls	13	—.14	29	.34	14	.61*
Total Group	113	.33**				

* $p < .05$.
** $p < .01$.

TABLE 4

MEANS OF THE CORRELATIONS BETWEEN RT AND MT FOR SUBJECTS OF EACH SEX AND THREE LEVELS OF ANXIETY

	LEVELS OF ANXIETY					
	LOW		MEDIUM		HIGH	
	N	*r*	*N*	*r*	*N*	*r*
Boys	13	.07	31	—.07	13	—.11
Girls	13	.02	29	—.04	14	—.01
Total Group	113	—.03				

the 113 correlations reached the .05 level of significance or higher, 14 in the positive direction and 24 in the negative direction. With an *N* of 40 it is necessary to have a correlation of approximately .31 at the .05 level.

One further estimate of RT-MT relationship was attempted by correlating the two scores for the 113 *Ss* on each of the 40 trials. The correlations ranged from —.10 to .42. Only eight of these correlations were greater than the value of .18 necessary to reach the .05 level of confidence. There was no apparent trend in the size of the correlations over trials.

DISCUSSION

The finding of no relation between anxiety and reaction speed, regardless of the measure used, supports earlier studies (1, 4, 7). These results do not support the hypothesis derived from Hullian theory (6) that increases in anxiety, or drive, lead to increases in response strength. Since there is a considerable body of evidence, recently summarized by Taylor (15), which supports Hull's theory, using the MAS as a measure of drive, these results would suggest that the reaction time task has some peculiarities not common to the other types of tasks where data have been obtained supporting the theory.

The most obvious difference is the measure of response strength used. For the most part, those studies supporting the theory have used number of correct or conditioned responses as measures of response strength. These are relatively gross measures in terms of the amount they are affected by competing responses aroused by irrelevant variables present in the experimental situation. It has been pointed out (11, 13), however, that speed measures are highly sensitive to competing responses aroused in the experimental situation. It would appear that in order to provide an adequate test of the theory, using speed as a measure of response strength, control of all variations in extraneous stimulation arousing competing responses would have to be instituted. Such control would be particularly difficult with human subjects for, as Castaneda (1) has suggested, the human subject may give himself verbal instructions which may compete with the response being measured in the reaction time task. In this connection, it should be noted that Castaneda found a significant relation between amplitude and MAS in the reaction time task although the reaction speed scores showed no significant relation to MAS. Presumably response amplitude is not as sensitive to competing responses.

One might question the CMAS as a measure of anxiety, or drive, in this study although other studies have shown this scale to have a similar relation to performance in learning situations as does the MAS (2, 9). An alternative hypothesis might be that the instructions to respond quickly, given by an adult to a child, may have maximized the general drive level and, therefore, minimized differences attributable to anxiety. In addition, the temporal restriction of responses measured in this task may not have allowed enough variability between subjects for differences to become apparent.

The correlation of .33 between the mean RT and MT speed scores confirms previous findings (10) and indicates that over a series of trials the two measures are related. The fact that the correlation is so low would suggest, however, that two different responses are involved and, therefore, ought to be treated separately in future studies of reaction time. Such is the case with animal studies where different relations frequently are obtained between various independent variables and starting time (response evocation time) and running time in a runway situation.

The average correlations for groups of *S*s revealed no relation between RT and MT; however, for individual *S*s there were a large number of correlations which reached a statistically significant level of confidence. This range of correlations, in both the positive and negative direction, would indicate that some variable may be operating to produce differential effects in the response patterns of the *S*s involved. It would appear that anxiety is not a variable affecting this phenomenon, but the factors which may affect the relation are not apparent on the basis of these data. The fact that the correlations between the two measures remain low over the 40 trials would indicate that the variables which may be involved tend to operate in a consistent manner over a relatively long series of trials.

SUMMARY

The present study was concerned with starting speed and movement speed in a reaction time task as related to scores on the CMAS. A total of 113 sixth grade *S*s, previously given the CMAS, were given 40 trials in a simple reaction time task. They were required to start with their finger on a microswitch and, upon presentation of a light, press another button. Time was measured from onset of light to release of microswitch and from release of microswitch to pressing the button.

Boys were significantly faster than girls, and there were significant increases in speed over trials on both measures; however, there were no significant differences attributable to anxiety on either measure. A low but significant correlation was found between the mean RT and MT speed scores for the total group. Correlations between the two measures over trials for individual subjects were distributed over a wide range. The correlations between the two measures for each trial showed no consistent trends over trials.

The results were discussed in terms of the sensitivity of the speed measures to competing responses and the advisability of using both RT and MT measures in future studies.

References

1. Castaneda, A. Reaction time and response amplitude as a function of anxiety and stimulus intensity. *J. abnorm. soc. Psychol.,* 1956, 53, 225-228.
2. Castaneda, A., McCandless, B. R., & Palermo, D. S. The children's form of the manifest anxiety scale. *Child Develpm.,* 1956, 27, 317-326.
3. Castaneda, A., Palermo, D. S., & McCandless, B. R. Complex learning and performance as a function of anxiety in children and task difficulty. *Child Develpm.,* 1956, 27, 327-332.
4. Farber, I. E., & Spence, K. W. Effects of anxiety, stress, and task variables on reaction time. *J. Pers.,* 1956, 25, 1-18.
5. Grice, G. R. Discrimination reaction time as a function of anxiety and intelligence. *J. abnorm. soc. Psychol.,* 1955, 50, 71-74.
6. Hull, C. L. *Principles of behavior.* Appleton-Century, 1943.
7. Kamin, L. J. Relations between discrimination, apparatus stress, and the Taylor scale. *J. abnorm. soc. Psychol.,* 1955, 51, 595-599.
8. Kamin, L. J., & Clark, J. W. The Taylor scale and reaction time. *J. abnorm. soc. Psychol.,* 1957, 54, 262-263.
9. Palermo, D. S., Castaneda, A., & McCandless, B. R. The relationship of anxiety in children to performance in a complex learning task. *Child Develpm.,* 1956, 27, 333-337.
10. Pierson, W. R. The relationship of movement time and reaction time from childhood to senility. *Res. Quart.,* 1959, 30, 227-231.
11. Postman, L., & Kaplan, H. L. Reaction time as a measure of retroactive inhibition. *J. exp. Psychol.,* 1947, 37, 136-145.
12. Rossi, A. M. An evaluation of the manifest anxiety scale by the use of electromyography. *J. exp. Psychol.,* 1959, 58, 64-69.

13. Spence, K. W. *Behavior theory and conditioning.* Yale Univer. Press, 1956.

14. Taylor, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.,* 1953, 48, 285-290.

15. Taylor, J. A. Drive theory and manifest anxiety. *Psychol. Bull.,* 1956, 53, 303-320.

16. Wenar, C. Reaction time as a function of manifest anxiety and stimulus intensity. *J. abnorm. soc. Psychol.,* 1954, 49, 335-340.

ROLE OF DELAY OF REWARD IN SPEED OF SIZE AND FORM DISCRIMINATION LEARNING IN CHILDHOOD[1,2]

GLENN TERRELL *and* ROBERT WARE*
University of Colorado

Numerous studies at the infrahuman level have shown that the strength of a habit is in part a function of the temporal nearness of the reinforcement to the response. Spence (6) has reviewed the literature related to this phenomenon. Perin (5) and, more recently, Logan (4) have shown that learning in the rat is related to the time interval separating reaction from reward. Perin found a negatively accelerated gradient of rate of learning which fell to zero at a delay of about 30 seconds. Logan demonstrated that the rat develops a preference for an act which is followed by a shorter delay of reinforcement.

Lipsitt and Castaneda (3) in an instrumental learning set-up found that young children show a preference for and respond more quickly to a stimulus which has been associated with immediate reinforcement as opposed to another stimulus which is associated with a reinforcement delay. Only one study in this area has been done with children in a discrimination learning arrangement. Terrell (7) found that children can learn a "larger-than" discrimination as quickly under delayed reward conditions as they can when immediately rewarded provided they are able to observe concrete evidence of progress toward the to-be-received reward under the delayed condition. Several factors, however, make this experiment somewhat difficult to interpret. It was contaminated to some extent by a magnitude of reward variable since *S*s assigned to the immediate reward condition received different amounts of reward, candy, due to wide individual differences in time required to learn. This did not apply to the *S*s of the delayed reward group since their reward, a small bag of candy, was delivered after they had learned the task. Also, all *S*s in this study received an immediate reward, a light flash, as an indication of a correct response. Finally, since the candy for both the immediate and delay group *S*s was administered by *E* in the aforementioned study, there is an apparent lack of precision

---

* Department of Psychology, University of Colorado, Boulder.

[1] This investigation was supported by a research grant, M-2063, from the National Institute of Mental Health, National Institutes of Health, United States Public Health Service. Grateful acknowledgment is due Nevin J. Platt, Melvin Wiesley, Keith Dillman, and all the teachers of the Martin Park School of Boulder, Colorado, without whose assistance this experiment could not have been performed.

[2] The writers wish to thank Charles C. Spiker who contributed several valuable ideas contained in this paper.

Child Develpm., 1961, 32, 409-415.

in the time interval between the *S*'s response and the reward. It is believed that these problems are remedied in the experiments described herein. Additionally, two tasks are given *S*s in each of the two studies reported below, a fact which increases the generality of the results.

The investigations described herein are ones in which the effectiveness of an immediate reward was compared with a delayed reward in the speed of size and form discrimination learning in childhood. The reward employed was a light flash. Two-stimulus problems were used in experiment I, while three-stimulus problems were used in experiment II.

METHOD

Subjects

The *S*s were 80 children. Twenty kindergartners and 20 first graders were used in each experiment.

Experiment I

Materials. All stimuli were three-dimensional geometric figures. The stimuli for experiment I were as follows: For task 1, two cube boxes having basal areas of 2 and 4 sq. in., respectively; for task 2, a sphere 3 in. in diameter and a pyramid 3 in. in height, 4 in. on each side, and 4 in. on the base. Additional apparatus consisted of a background board and a panel board. The background board was a 13½ by 13 in. piece of plywood with two push-button mounts which were located 2 in. from the right and left edges of the background board and 1¾ in. from the front edge of the board. On each trial the stimuli were placed onto the background board immediately in front of the push-button mounts, A piece of terry cloth was glued to the board where the stimuli were placed in order that the sphere would not roll out of position. Batteries and circuits necessary to operate a signal light were concealed under the background board. Locked onto the rear edge of the background board was a 7⅞ by 10¾ in. panel board which contained the signal light. This light was mounted in the center of the panel board, 1⅛ in. from the top. The circuits were so arranged that a correct response caused the light to go on. The light went on immediately for the immediate reward condition, while there was a delay of 7 sec. for the delayed reward condition. Two Hunter decade type timers controlled the delay of reward and the 1-sec. duration of the light flash.

Design and procedure. There were two categories each of two experimental variables. One variable was the reward condition, while the other was the task factor. The two reward categories were: (a) an immediate light flash after a correct response and (b) a 7-sec. delay in the light flash following a correct response. The two tasks were: (a) a button-pushing response to the larger of two cube-boxes and (b) a button-pushing response to a sphere instead of a pyramid.

*S*s were randomly assigned to two groups of 20 children each. Group I *S*s were given the size task under the delayed reward condition and the form task under the immediate reward condition. Group II *S*s were given the size task under the immediate reward condition and the form problem with a delayed reward. With this type of design, a Lindquist Type II arrangement (2), all *S*s were administered both conditions of each experimental variable, thus making it possible to control for individual differences in evaluating the effects of both the reward and task variables.

*S*s of both groups learned the two tasks concurrently. That is, from trial to trial the order of administering the tasks, along with the appropriate reward condition, was counterbalanced according to the following sequence: Size, form, form, size, size, form, size, form, form, size. This procedure minimized contamination of the incentive and task treatments with each other due to order effects. The position of the positive stimulus was randomized alike for each *S*.

The *S*s were tested individually. Each *S* received the following instructions: "This is a game where we play with these things (*E* shows the two cube-boxes and the sphere and pyramid to the *S*). Sometimes these two (*E* picks up the two cube-boxes) will be put here (*E* places the cube-boxes on the background board immediately in front of the push buttons) and sometimes these two (*E* picks up the sphere and pyramid) will be put here (*E* places the sphere and pyramid on the background board). You are supposed to choose one of these (*E* again points to the cube-boxes and the sphere and pyramid) and push the button (*E* points to the push buttons) in front of the one you choose. If you are right, that light will go on (*E* points to the light on the panel board). If you are wrong, the light will not go on. Now remember, the game is to see how soon you can make the light go on every time." The last sentence was repeated after the fifth trial in a further effort to insure that the *S*s understood the point of the task. Training was continued until *S*s were correct on six consecutive trials.

Although the main dependent variable of interest was the number of trials to criterion, it was felt it would be a simple matter to test for the effects of immediate and delayed reward on the *S*'s preferences for the two sets of stimuli. Accordingly, just prior to the first trial, and again after the last trial, all *S*s were asked to express a preference for the size or form stimulus sets. This measure was obtained by simply asking the *S*, "Which of these do you like better—these two (*E* points to cube-boxes) or these two?" (*E* points to the sphere and the pyramid).

Experiment II

Experiment II was conducted primarily for the purpose of extending the comparison of the delay of reward variable to tasks assumed to be more difficult than those in experiment I. Thus, experiment II provides for a test of the generality of the treatment effect.

The *Ss*, materials, experimental groups, and design for experiment II were the same as for experiment I with the following exceptions: (a) A different group of 40 *Ss* was used in experiment I. (b) An additional push button was installed midway between the two used in experiment I. (c) The stimuli for experiment II were, for task 1, three cube-boxes having basal areas of 2, 3, and 4.48 sq. in. (d) The correct response for task 1 was a button-pushing response to the largest of the three cube-boxes. (e) The stimuli for task 2 were a cylinder, sphere and a pyramid, all 3 in. in height. (f) The correct response for task 2 was a button-pushing response to the pyramid. (g) In the instructions to the *S* the *E* demonstrated the stimuli appropriate to the three stimulus size and form tasks instead of those for the two stimulus problems of experiment I. And finally, (h) the training was continued until *S* made four consecutively correct responses.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF TRIALS TO CRITERION, ORIGINAL DATA
(Each total treatment group, $N = 40$)

	IMMEDIATE		DELAY		TASK TOTALS	
Task	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experiment I						
Size	6.9	6.1	15.1	14.7	11.0	12.0
Form	6.4	6.1	18.1	13.2	12.2	11.8
Treatment Totals ..	6.6	6.1	16.6	14.0		
Experiment II						
Size	10.7	15.7	27.7	14.2	19.2	17.2
Form	21.9	17.0	17.8	15.0	19.9	16.2
Treatment Totals ..	16.3	7.0	22.7	15.5		

RESULTS

In both experiments an analysis of variance was made on the number of trials to the criterion. Also calculated was a χ^2 test for significance of change in *S*'s preference of the stimulus sets as a function of reward condition. Because of data reported by Boneau (1) it was felt that there was no value to making the usual tests of homogeneity and normality. Boneau reports that particularly when sample *Ns* are equal and more than 30 the *t* test will result in probability figures that are highly accurate even though the normality and homogeneity assumptions may not be met.

Table 1 contains the acquisition data for both experiments. The summary of the variance analysis for experiment I appears in Table 2, for experiment II in Table 3. As can be seen in Table 1, *Ss* learned significantly more

TABLE 2

ANALYSIS OF VARIANCE OF NUMBER OF TRIALS TO CRITERION
EXPERIMENT I

Source	*df*	*MS*	*F*
Between Subjects	39		
Rewards × Tasks	1	61.25	.38
Error (b)	38	161.34	
Within Subjects	40		
Rewards	1	2,000.00	24.24***
Tasks	1	31.25	.38
Error (w)	38	82.52	

*** $p < .001$.

quickly in both experiments when rewarded immediately than when given a delayed reward. In experiment I the means were 6.6 and 16.6 for the immediate and delay conditions, respectively, ($p < .001$). In experiment II the treatment means were 16.3 and 22.7 for the immediate and delay conditions, respectively, ($p < .001$). The interaction and task effects were nonsignificant in experiment I. The task effects were nonsignificant, but the interaction was significant ($p < .05$) in experiment II.

The prediction in the test of significance of change in *S*'s stimulus set preferences was that preferences would change in the direction of the set with which immediate reward had been associated. In both experiments the changes were in the predicted direction but nonsignificantly so ($p = .35$) for both experiments.

TABLE 3

ANALYSIS OF VARIANCE OF NUMBER OF TRIALS TO CRITERION
EXPERIMENT II

Source	*df*	*MS*	*F*
Between Subjects	39		
Reward × Tasks	1	2,215.51	4.77*
Error (b)	38	464.42	
Within Subjects	40		
Rewards	1	825.61	19.11***
Tasks	1	9.11	.21
Error (w)	38	43.20	

* $p < .05$.
*** $p < .001$.

Discussion

In a convincing manner the results of these experiments in general support the delay of reinforcement studies performed at the infrahuman level and the Lipsitt and Castaneda experiment with children. In both experiments herein reported the *S*s learned more effectively when given an immediate reward than when rewarded following a 7-sec. delay. Contrary to the results obtained by Lipsitt and Castaneda, however, the *S*s of the present studies did not show a significant preference at the end of the experiment for the stimulus set with which immediate reward had been associated. In view of the fact that the *S*s of the Lipsitt and Castaneda experiment received 60 trials, it would appear that in the present study the number of trials was insufficient to create a substantial difference in *S*s' preferences. It appears then that preference measures are less sensitive to the time of delay of reward or are determined by more random variables in the past history of the *S*s than are the acquisition measures.

The interaction which was found in experiment II deserves some consideration. An examination of Table 1 reveals that, in experiment II, *S*s learned the size task more quickly under the immediate reward than under the delayed reward, but that the reverse of this was true with the form task. Two observations seem crucial here. First of all, the *S*s of group II generally were more adept at learning both tasks than those of group I, so that it is possible that the effects of the delayed reward on group II when learning the form task were mitigated by this factor. The mean number of trials for both tasks for group II was 14.3 while that of group I was 24.8 ($p <$ .01). The second observation of seemingly considerable importance in interpreting this interaction is that, while the difference between the size-immediate and the size-delay treatment combinations is significant at the 1 per cent level in the predicted direction, the form-immediate and the form-delay difference is nonsignificant. In other words, with respect to neither of the tasks learned in the two experiments was learning quicker under the delay condition. For these reasons the significant interaction of experiment II seems rather unimportant.

The question of why an immediate reward should result in more efficient discrimination learning than a delayed reward must be dealt with. Is the associative (H) factor being effected by the time interval between response and reinforcement, or is it the incentive (K) factor? There are suggestions, however casual, that perhaps both *H* and *K* are being effected by this time interval. *S*s apparently responded more quickly, i.e., made their choice and pressed the button more rapidly, when learning under the immediate reward than when learning under the delayed condition, a fact which suggests that *S*s were more certain of their choice being the correct one when rewarded immediately. On the other hand, the incentive (K) factor apparently was operating also in that *S*s seemed to lose interest more readily when learning under the delay condition. The loss of interest referred to consisted of *Ss* gazing about the room, asking random questions, fidgeting, and exhibiting in general a decreased vigor of button-pushing responses. Research designed to test these ideas is currently in progress.

SUMMARY

This paper reports the relative effects of a delayed reward and an immediate reward (light flash) in two experiments involving size and form discrimination learning problems. Forty *Ss*, kindergartners and first graders, were used in each experiment. In both experiments, *Ss* learned significantly more quickly when rewarded immediately following correct responses than when rewarded following a delay of seven seconds. There were no task differences. It is believed that both associative and incentive factors played a role in determining speed of learning.

REFERENCES

1. BONEAU, C. A. The effects of violations of assumptions underlying the *t* test. *Psychol. Bull.,* 1960, 57, 49-64.
2. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Houghton-Mifflin, 1953.
3. LIPSITT, L., & CASTANEDA, A. Effects of delayed reward on choice behavior and response speeds in children. *J. comp. physiol. Psychol.,* 1958, 51, 65-67.
4. LOGAN, F. A. The role of delay of reinforcement in determining reaction potential. *J. exp. Psychol.,* 1952, 43, 393-399.
5. PERIN, C. T. The effects of delayed reinforcement upon the differentiation of bar responses in white rats. *J. exp. Psychol.,* 1943, 32, 95-109.
6. SPENCE, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.,* 1947, 54, 1-8.
7. TERRELL, G., JR. The role of incentive in discrimination learning in children. *Child Develpm.,* 1958, 29, 231-236.

A HUMAN BIOLOGIST LOOKS AT THE CONTEMPORARY SCENE[1]

WILTON MARION KROGMAN*
University of Pennsylvania

At the risk of overemphasis I'd like to observe that biology is the fact of life, culture the factor of living; biology is the substance of life, culture the form of living. Biology is thus the very nucleus and matrix of all that is. All else is peripheral. Biology provides the force and direction to the stream of life. All else provides environmental background of living and existence.

As a human biologist I'd like to observe that we cannot evaluate the contemporary scene as an achieved isolate. What *is* happening to Man *has* happened to Man. What we contemplate Today has its roots in Yesterday. Man is an evolving, not an evolved, creature, and, I am constrained to say, this is in principle even more true of what he has wrought in terms of cultural complexities. We reap a whirlwind Now because a wind was sown Then!

What I propose to do it to limn our problems in the perspective of evolutionary time, for Man is one end-result of eons of emergent development. We begin, simply, with a ganglion—an aggregation of nerve cells—at the anterior end of an ancient form of life. This ganglion, this raw knot of action and response cells, was the forerunner of the *brain* which was to become the guiding agent first in more or less random response, then in coordinated reflex interaction, and finally in conscious and directed behavioral reactions dictated by a cortex that had evolved into a *mind*. Now the brain no longer runs a body alone, it directs and guides a life! From mere neurophysiology has arisen the incredible and tortuous complexity of *social behavior*. To the organic has been added the supraorganic!

Out of all of the many facets of the current cultural scene I'd like to focus, briefly, upon two: human relationships and human numbers, both, in my opinion, biologically determined, but culturally oriented and mediated.

---

* Graduate School of Medicine, University of Pennsylvania, Philadelphia.

[1] Address of the President at the Biennial Meeting of the Society for Research in Child Development, Pennsylvania State University, March 17, 1961.

Child Develpm., 1961, 32, 417-422.

I refer, of course, to problems of race and to problems of population composition. I shall consider them in this order.

There is a major premise to be accepted at once: all people now living, or who have lived for the past 100,000 years or so, belong to the same genus, *Homo,* and to the same species, *sapiens*. To make the identity even more graphic, it may be observed that 29,500 pairs of the 30,000 pairs of human genes are shared equally by all of Mankind. It is the remaining 500, withdrawn differentially from the common gene pool, that set up so-called "racial differences"—those external hallmarks of skin, hair, nose that grant measures of unlikeness. In the broadest possible terms there is but one "human race."

Notwithstanding the foregoing generalization, there *are* subdivisions called races. What, then, is a race? Genetically speaking, a race is a breeding population within *Homo sapiens* with a certain combination of heritable traits to give a degree of physical distinctiveness. The population unit is subject to a certain measure of isolation, either geographical (formerly) or culturally (currently).

How many races are there? It depends upon the criteria, or combination thereof, set up for classification. There are "macrogeographical" and "microgeographical" races. Hence, the number may vary from three or four to 30. Today we generally accept eight major geographical races: European, African, Asiatic, (East) Indian, Australian, Melanesian, Polynesian, and Amerindian.

Our concern is not really with either definition or classification of race: it is with cultural *reaction* to the biological fact of race. The problem is not one of whether or not a person looks differently, but the *assumption of difference*. There are four pretty well-defined reactions to and beliefs about, race differences: (a) they are fixed, immutable, unassimilable; (b) so-called racial traits (behavioral, intellectual) are innate, i.e., genetically entrenched; (c) the greater the degree of *seen* difference, the greater the separation in a superior-inferior hierarchy; (d) hybridization ("miscegenation") can only result in biological inferiority or deterioration ("the lower will drag the upper down"). These are the ghosts which must be laid before "race problems" cease to haunt us.

In 1950 the population of the United States was composed as follows:

Native white	124,780,860
Foreign-born white	10,161,168
Negro	15,042,286
Indian	343,610
Japanese	141,768
Chinese	117,629
Other	110,240

In a sense this breakdown is really ludicrous. There are no "native" whites; the only true "natives" are the Amerindians. Be that as it may, we Americans are neither biologically nor culturally homogeneous. There *are* differences among us: in appearance, in cultural level, in performance, in behavior and attitude, in achievement, and in ability and potential. These are an observable and measurable set of facts. The problem is whether or not these variable levels are race-linked. Is there a variability, in a given direction, that is innate? Is variability interracial, or is it intraracial? For racial friction do we blame *Biology* or do we penalize *Culture*? What is behind a 1944 report that we Americans ranked peoples in order of most to least preferred: white American, Canadian, English, French and Swiss, Spanish and Portuguese, Jew, Negro, Turk, Chinese? Why such a cultural hierarchy?

In purely biological terms there is no validity to claims of "superior" racial status. As we move from a solely structural consideration of race, based upon morphology, to a functional analysis of race, based upon physiology and biochemistry, we are beginning to realize that there are certain adaptive traits that possibly confer greater survival values, with specific reference to the ecological niche occupied at the time the type was evolving (being selected). That is to say, for example, that a pigmented skin is better adapted (hence, "superior") to a tropical sun; or a prominent malar complex with fat pads to a sub-Arctic or Arctic climate.

It is time to quit calling a spade an agricultural implement. It is time to come to grips not with race, but with *racism*. It is time to face up to the fact of *prejudice*: the creation of advantages and material benefits for the prejudicor, the limitation or denial of these to the prejudicee.

We are making progress. The FEPC, antidiscrimination decrees in the armed forces, and desegregation in the schools are all ameliorative. But we continue to pay a high sociocultural price for barriers of prejudice. In our economy prejudice has caused an unbalanced use of manpower which, in turn, has set up a maze of welfare problems. In broader perspective prejudice has created a national and political schism and aggravated regional tensions. Prejudice is a hindrance to free and clear sociopolitical communication, both at national and international level. It engenders a frustration-aggression pattern of social interaction, and it warps the attitude and personality (group and individual) of those prejudiced. Prejudice feeds on itself and is contagious: it is an easy progression from anti-Negro to anti-Jew, or anti-Catholic, or anti-upper class, or what you will.

I repeat, we are making progress. But, I have heard it said, "knowledge is slow of foot, and wisdom limps far behind knowledge." In current terms we are witnessing a lag between legislative and judicial action and sociocultural implementation. In oversimplified terms the issue of school desegregation is a conflict between law and custom. Since law represents the codefied will of the majority, there can be no doubt of either its validity or its ultimate and absolute triumph.

I think that we are all coming to realize that problems of race relations are generic: the Negro-white problem is really one with the minority-dominant problems of Catholics, Jews, Spanish-Americans, Japanese-Americans, and so on. I feel strongly that we Americans can resolve our so-called race problem only if we come to the realization that such a problem is little or no different from any other intergroup or interpersonal problem in our everyday cultural life. I have every hope that experience, reason, understanding, and a give-and-take attitude will triumph over prejudice and its accompanying complex of ignorance, fear, and emotion.

Here is where we, people of goodwill working with children, can do a lot. It is said that race prejudice must be learned. We can, by precept and by admonition, help to see that it is either not learned or unlearned. The receptive mind of the child will store ideas, concepts, and examples of "fair play," of equality of opportunity, of sharing the good things of the American way of life. Democracy, full and even participation of the everydayness of living, is not a race-limited or a part-time affair.

I want to turn now to an even more critical problem. It, too, is biological: what *can* be more biological than breeding? I refer, of course, to the aptly-termed "population explosion."

Once more I hold to the long-term view. In his earlier days *Homo* was irregularly and sparsely spread over Eurasia. Not until the end of the Ice Age, some 25,000 years ago, did he begin his deployment. Since that time two great cultural events have spurred his increase and his spread: (a) the "Agricultural Revolution" of about 10,000 years ago; (b) the "Industrial Revolution" (or "Scientific Revolution") of about 1650 A.D. Man in the Old Stone Age may have numbered 10 million; in 6000 B.C. he could count 20 million; by 1200 B.C., 100 million; at the opening of the Christian Era, 350 million; by 1650 A.D. varying estimates run from 470 million to one billion. (In Europe, alone, the 1650 to 1950 A.D. increase was from 100 to 900 million).

In more realistic terms we may summarize as follows: Man achieved his first billion by 1830; the second by 1930 (100 years elapse); his third by 1965 (35 years elapse); his fourth by 1980 (15 years elapse); his fifth by 1990 (10 years elapse); his sixth by 2000 (10 years elapse). To drive this closer home, the United States will add 100 million by 1980, to bring us to 280 million. One of every 20 men who have ever lived is alive today. The time is not too far away when a tenth of all humans who have ever lived will walk the Earth simultaneously! If your pulse ran 100 to the minute, every pulse beat would record a birth!

In the United States we are facing a problem of urbanization. In 1950 about 57 per cent of the United States population lived in 168 metropolitan areas (cities of 50,000 plus). From 1900 to 1950 such areas absorbed 73 per cent of total population growth; in the last decade of this period the percentage rose to 81. By 1975 we shall have 200 such metropolitan areas.

Over and beyond the urban trend is the likelihood of uneven geographic distribution in the United States. By areas the percentage growth for 1955 to 1970 may be calculated as follows: (a) 51 per cent for the three Pacific

States; (b) 44 per cent for the eight Mountain States; (c) 32 per cent for the five East-North-Central States; (d) 28 per cent for the eight South Atlantic States; (e) 22 per cent for the three Mid-Atlantic States; (f) 21 per cent for the four West-South-Central States; (g) 16 per cent for the seven West-North-Central States; (h) 16 per cent for the six New England States; and (i) 5 per cent for the four East-South-Central States.

Still being ethnocentric and looking at the United States, what are some of the consequences? Let me cite a few quite at random. The 65 million cars now on the highways if placed bumper-to-bumper 10 abreast on a 10-lane highway would encircle the Equator. Everytime we build a mile of superhighway we lose 40 acres of land. In a little over 10 years we will have dumped 60 million gallons of radioactive residue, with a life-span of from a decade to 24,000 years. Millions of Americans now drink treated sewage. In 1919 to 1956 the United States population increased 73 per cent, but the number on the government payroll went up 166 per cent. In 1912 a Senator represented 993,000 people, a Congressman 219,000; in 1960 the numbers were, respectively, 1,800,000 and 400,000; by 1975 they will be 2,225,000 and 517,000. By 1965 one half of all children born in New York City will be born to indigent families. I could go on and on with data such as these; some of them almost seem to be *non sequiturs*, but they are all tied in, one way or another, with more and more people—and still more!

In an over-all view the problem of numbers resolves to two consideraations: *reduce them* or *feed them*. The first makes more sense to me, with a full realization of the ethical and moral issues. Controlled, i.e., limited, breeding is the only solution: world-wide contraception. The real issue centers around the ultimate morality of underprivileged conditions for one half of a tremendous population or more evenly distributed cultural and material benefits for a smaller population. I think we must squarely face the issue: have we the moral right to subordinate *quality* of population to *quantity*? It reduces to this: do we favor a lot of "have-nots" or fewer "haves"? When I refer to quality vs. quantity, have-not vs. have, I am referring to *both* biology and culture. I include both materialistic and spiritual values. The squalor and debasing circumstances of rank overpopulation must leave their imprint upon body and spirit alike. The dignity of complete individuality must not be swallowed in the morass of overwhelming numbers.

The land area of the Earth is about 55 million square miles, or about 36 billion acres; only about 2 per cent is readily arable (though Man has stretched this to 10 per cent). If our nearly three billion population were evenly distributed, there would be 40 people per square mile. Yet 66 per cent of the three billion live on only 8 per cent of the land surface. The end result? Anywhere from 66 to 75 per cent of the three billion live on a subminimal diet; they go to bed hungry every night. If we are to remedy this, the food supply will have to be doubled within the next generation (say, 25 years). Answers are available, in potential at least: supplementary irrigation; more extensive use of fertilizers; plant breeding for specific environment and for insect and disease resistance; the use of special chemical media, hydroponics, for example. If the foregoing could be completely and integratively developed, we could hope for a 25-fold yield increase, which could, in theory, support a total population of 50 billion.

There is a very important corollary of *absolute* numbers, and that is the problem of *relative* numbers. Not only are more people living, but they are living longer! By 1800 A.D., after nearly 2000 years of post-Roman history, Man achieved a life-expectancy of 35 years. From 1800 to 1950 this was *doubled* to 70 years. This is to say, we have made more gain in the past 150 years than in all recorded and prerecorded human history.

Man now is at a stage where he, alone of all evolved life, may direct and influence his future evolution. It is not so much of the body that I speak, for I feel it will not change too much in the next 100,000 years or so (unless, of course, radiation does open the Pandora's box of unpredictable mutations). It is the mind, the spirit, the very soul, of Man that will be focal in his self-mediated evolutionary path. Man's value-needs will guide the emergent human being. It has been said that Man must have a "feeling of certitude," that the Darwinian "struggle for existence" has given way to the (perhaps Freudian?) "struggle for fulfillment." In essence this cuts across all human problems of race, of numbers, of age-levels, for here is implicit the essential value of the human individual. At this level fulfillment means the facilitation of his own personality, the realization of his own innate and learned capacities, the satisfaction of material and spiritual needs and drives, and the development throughout life of new qualities of experience. Here, for each human being, there should be no barriers of race, color, creed, socioeconomic status, or age-level!

I envision Man—*all men*—as sharing equally the greatness of what *Homo sapiens* has wrought: Man makes conscious choices, governed by a conscience; Man works with abstractions and conceptual frameworks, so that right and wrong, in both abstract and concrete, may be adjudged; Man formulates his own ultimate goals and establishes the values to be achieved; Man has defined and created beauty for himself and set in motion the mechanisms for its realization and expression; and Man has formulated the moral and ethical qualities of evil vs. good, with guilt the arbiter or penalty of choice or evaluation—Man's the freedom to choose and to pursue.

I conclude what I have to say with an affirmation of faith not alone in Man but in God. If in a sense Man has walked through the ages in accordance with natural law, he will continue to do so. Yet, I would rephrase it in the words of the Psalmist of old who admonished that Man "walk humbly" with his God. In this sense it is Man's destiny to stride confidently into the future with a renewed and deepened faith in the God of all life.

AN INVESTIGATION OF MATERNAL CONCERNS BY MAIL SURVEY

William O. Robertson*
The Children's Hospital, Columbus, Ohio

In recent years, pediatric attention has been increasingly directed toward well child supervision. Prevention rather than treatment of "disease" is the ultimate objective. An important facet of the program that has evolved is the pediatrician's function in advising and reassuring mothers about the common concerns of child rearing.

Surprisingly little is to be found in the literature about the focus or incidence of these common concerns. Admittedly, many factors influence the development of maternal concern, its transmission to the physician, and his perception and interpretation of it. The complexity of these influences and the costs of investigations that might control each variable effectively make it understandable that a paucity of objective data on the subject exists. Nevertheless, if the "New Pediatrics" (5) is to progress, reliable norms about this as well as other aspects of maternal behavior would appear essential.

The mail questionnaire method of survey has established its value in various nonmedical fields. Currently it is also being employed effectively in medical research; for example, in longitudinal studies on the incidence of cigarette smoking (10) and follow-up studies on the dissemination of pathogenic staphylococci from newborn nurseries (9). Consequently, it seemed worthwhile to explore the usefulness of this technique in securing childrearing data. Therefore, a mail questionnaire survey directed to mothers of young infants was undertaken in March, 1957. To determine, among other things, the consistency of its findings, it was repeated in November, 1958.

METHODOLOGY

A mail questionnaire listing various sources of potential maternal concern was developed. Explanatory instructions requested that the mother examine each category and note her answer simply by checking the appropriate response. Space was provided for her to describe or explain the nature of any existing disease or abnormality, as well as her physician's recommended

---

* The Children's Hospital, Columbus 5, Ohio.

Child Develpm., 1961, 32, 423-436.

therapy, and to add comment she might consider appropriate. Certain specific items, such as the infant's sex, the maternal age, etc., were also requested, to provide data for distributional analysis and correlative purposes.

Preliminary trials were undertaken to test adequacy of instructions, comprehension of terminology, and degree of response. Necessary modifications were then incorporated into the questionnaire and its final format prepared, as shown in the Appendix.

The names of 2,000 mothers who had delivered during a specified interval were obtained from existing commercial sources. These names are alleged to have been assembled in a random fashion from a master list said to include 85 to 90 per cent of the nation's births. Analysis of their geographical distribution showed them to be statistically representative.

A letter, signed by the author, was then forwarded, together with a questionnaire, to each name listed. It congratulated the mother and requested that she complete and return, anonymously, the enclosed questionnaire in an accompanying prepaid addressed envelope. One week later, a second mailing of the letter, questionnaire, and envelope, supplemented with a notation thanking the mother if she had already returned the questionnaire, and requesting her to do so if she had not, was dispatched.

Replies on returned questionnaires were coded and entered on standard IBM cards. The geographical location of each respondent was determined from the postal marking on the returned envelope. These locations were assigned their specific county sites and appropriate population figures determined from available government statistics (12). Arbitrarily, those counties whose populations exceeded 100,000 or which bordered large metropolitan centers were designated as "urban"; otherwise they were "rural." Appropriate designations of "urban" and "rural" were entered on the individual IBM cards.

The distribution of the group responding was compared to control data with regard to regional geographical location, age, education, and parity of the mother, as well as sex, age, and birth weight of the infant (11, 12, 13). Incidences of different concerns were computed for specified variables.

Response

In this pilot investigation, two independent surveys were conducted, utilizing an identical questionnaire. The first, Survey A, was initiated in March, 1957. Of 2,000 mailed questionnaires 125 could not be delivered because of the addressee's temporary or permanent change of address or incomplete address. These were returned unopened. Of the 1,875 delivered questionnaires, 1,132 (60.5 per cent) were completed and returned. Thirty-nine were discarded because of (a) grossly inadequate completion of the form, (b) illegible postmarking on the return envelope, or (c) late completion or return of the questionnaire. A total of 1,093 returns was coded and entered on IBM cards for processing. The second survey, Survey B,

TABLE I

INCIDENCE OF MATERNAL CONCERN

	PERCENTAGE OF MOTHERS INDICATING						
	"Some Worry"		"Considerable Worry"		TOTAL WORRIES		
General Category	*Survey A*	*Survey B*	*Survey A*	*Survey B*	*Survey A*	*Survey B*	*Survey A & B*
1. *Digestion* (spitting, burping, gas, vomiting, colic, etc.)	34.7	33.7	7.7	7.6	42.3	41.4	41.9
2. *Breathing* (uneven, hiccoughs, gags, chokes, gasps, grunts, etc.)	34.8	32.2	4.2	3.8	39.0	36.0	37.5
3. *Bowel Movements* (odor, color, too often, too loose, hard, etc.)	30.9	29.0	7.4	7.4	38.2	36.4	37.3
4. *Skin* (oily, dry, rash, scratches, etc.)	29.2	28.2	3.7	3.3	32.9	31.4	32.1
5. *Sleeping* (too much, not enough, not regular, restless, etc.)	27.1	24.9	4.9	4.2	32.0	29.1	30.6
6. *Eating* (not enough, too much, not regular, hungry, disagrees, etc.)	25.3	22.5	5.2	4.0	30.5	26.5	28.5*
7. *Navel* (swollen, too large, small, bleeding, odor, etc.)	23.3	19.6	4.8	5.1	28.1	24.7	26.4
8. *Buttocks* (diaper rash, sore, color, etc.)	20.0	18.8	3.0	4.7	23.0	23.5	23.2
9. *Other* (spoiling baby, food preparation, bathing, clothing, diapering, etc.)	22.0	19.0	2.7	2.4	24.7	21.3	23.0
10. *Crying* (too much, too little, strong, weak, turn color, etc.)	17.3	14.7	4.3	3.4	21.7	18.1	19.9*
11. *Eyes* (puffy, red, crossed, color, etc.)	17.3	14.6	1.8	2.1	19.1	16.7	17.9
12. *Accidents* (while sleeping, eating, playing, bathing, etc.)	15.9	11.9	2.4	2.4	18.3	14.3	16.3*
13. *Head* (size, shape, soft spot, etc.)	12.5	10.9	2.5	1.7	15.0	12.6	13.8
14. *Mouth-Lips* (size, shape, color, sore, swallowing, thumbsucking, etc.)	7.8	9.5	1.4	1.8	9.1	11.2	10.1
15. *Weight* (not gaining, too fat, too thin, etc.)	7.1	6.9	0.6	1.7	7.6	8.6	8.1
16. *Nose* (size, shape, running, etc.)	8.4	6.0	0.6	1.0	8.9	7.0	8.0
17. *Legs-Feet* (too thin, too heavy, not straight, etc.)	5.5	5.1	1.7	1.2	7.2	6.4	6.8
18. *Urine* (odor, color, too often, too little, etc.)	4.4	6.9	0.6	.8	5.0	7.7	6.3*
19. *Ears* (shape, size, color, etc.)	4.7	5.1	0.5	.7	5.2	5.8	5.5
20. *Stomach* (too large, small, hard, soft, swollen, etc.)	4.5	4.0	0.5	.7	5.0	4.7	4.8
21. *Hair* (too much, too little, falling out, etc.)	3.5	3.7	0.6	.2	4.1	3.9	4.0
22. *Arms-Hands* (too thin, too heavy, not straight, etc.)	1.0	1.2	0.2	.3	1.2	1.4	1.3

* Differences between results of Survey A and B significant at .05 level but not .01 level.

was undertaken in November, 1958. Of 2,500 questionnaires distributed, 2,398 were delivered; 1,134 (47.5 per cent) were completed, returned, and analyzed.

When experimental data on age, parity, and level of education of the mothers, sex and birth weight of the infants, and geographical location of the families were compared with known data, no statistically significant differences were found except that, as was anticipated, the level of education of the respondents exceeded that found among women of comparable age groups in the general population. The age of the infants at the time the mothers completed the questionnaires ranged from 5 to 10 weeks; the mean age was 7 weeks.

Results

The frequencies with which different sources of concern were noted in the two surveys are listed in Table 1. Comparison of their findings shows that the incidence of "total worries" is significantly different at the 5 per

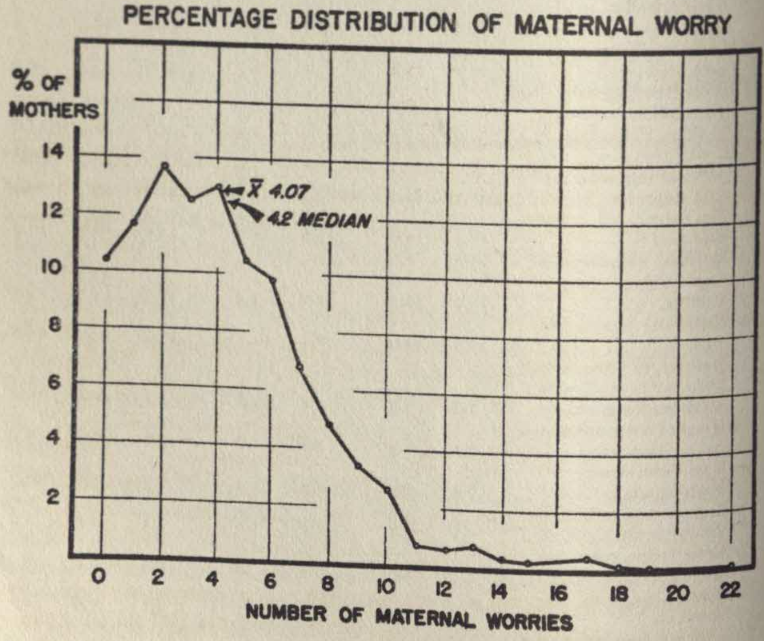

Figure 1—A distribution curve depicting the percentage of respondents checking none through the 22 items listed in the questionnaire. The mean number of items noted is 4.1; the median is 4.2. Of the mothers 10.4 per cent failed to note any item as a source of worry; only 5.3 per cent checked 10 or more items.

cent level in 4 of the 22 categories; at the 1 per cent level in none. Their respective rankings do not differ.

Although variations in, and overlap of, definitions must be considered, the categories designated "digestion," "breathing," and "bowel movements" were noted most commonly as sources of concern. Forty-two per cent of all respondents indicated some degree of worry about "digestion"; only 1 per cent, about "arms and hands." The two categories indicated most frequently as sources of "considerable worry" both related to function or malfunction of the infant's gastrointestinal tract.

A frequency distribution curve of the results from Survey A is shown in Figure 1. The average respondent checked 3.6 items as causes of "some worry" and .5 items of "considerable worry" for a total of 4.1 "worries" per mother.

TABLE 2

COMPARISON OF RESPONSE TO FIRST AND SECOND MAILINGS OF SURVEY A

	N	*Mean Number of Worries*	*SD*
1st Mailing	841	4.15	2.87
2nd Mailing	251	3.82	2.98

The mean numbers of concerns elicited in the first and the second mailings of Survey A are compared in Table 2; those from Survey A and Survey B are compared in Table 3. No statistically significant differences were found.

TABLE 3

COMPARISON OF RESPONSE TO SURVEY A AND SURVEY B

	N	*Mean Number of Worries*	*SD*
Survey A	1092	4.07	3.06
Survey B	1134	4.03	3.14

The frequency with which a specific "sign" or "symptom" was indicated as a source of concern in Survey A is noted in Table 4. The predominance of concern about functional or physiologic manifestations, as opposed to anatomic ones, is apparent. To be noted is that the fifth most frequently noted cause of worry was "spoiling the baby"; it was the second most frequently noted specific concern among primiparas.

Table 4

INCIDENCE OF SPECIFIC MATERNAL CONCERNS

(N = 1093 Mothers)

Specific Concern	*Percentage of Mothers Indicating Worry*	*Specific Concern*	*Percentage of Mothers Indicating Worry*
Skin Rash	23.4	Hungry Baby	9.1
Hiccoughs	17.4	Irregular Eating	8.9
Gas	16.7	Vomiting	8.8
Diaper Rash	16.1	Burping	8.6
Spoiling Baby	16.0	Swollen, Protruding Navel	8.4
Not Enough Sleep (Baby)	12.5	Infrequent, Painful BM	7.9
Too Much Crying	12.1	Choking	7.3
Irregular Sleep	11.9	Accidents While Asleep	7.2
Hard Stools	11.9	Frequent BM	7.0
Spitting Up	11.7	Thumbsucking	6.8
Colic	11.5	Dry Skin	6.8
Restless	11.1	Uneven Breathing	6.7
Loose Stools	10.9	Red Eyes	6.6
Bleeding Navel	9.4	Gags	6.3

To detect the influence that maternal age, uninfluenced by parity, might have in altering the notation of concern, all primiparous respondents were divided into three groups: less than 25 years of age, 25 to 29, and 30 and above. No significant differences were found between the mean numbers of concerns in these three groups.

To determine the influence that level of education might have in altering the mean number of concerns, all primiparous women under 25 years of age were grouped according to level of education; age and parity were thus held constant. No significant difference in concern was found between these groups.

To test for the influence of parity, those mothers 25 to 29 years of age who had completed some, but no more than a high school education, were grouped according to parity as shown in Table 5. As a group, mothers of first children indicated significantly more sources of concerns than did mothers of four or more children. Mothers of two children indicated significantly more concerns than did mothers of three or more children. In support of this finding is the fact that 13 per cent of responding multiparas spontaneously noted under "Comments" words to the effect that their concern had been much greater—both qualitatively and quantitatively—with their firstborn child.

Relatively large and statistically significant decreases in the incidence of certain concerns accompanied increases in parity. Such included "breath-

TABLE 5

RELATION OF PARITY TO MEAN NUMBER OF CONCERNS NOTED BY MOTHERS 25 TO 29 YEARS OF AGE WITH SOME OR COMPLETE HIGH SCHOOL EDUCATION

No. of Children	*N*	*Mean*	*SD*
1	51	4.35*	± 2.64
2	120	4.40**	± 2.98
3	121	3.69	± 2.70
4	80	3.20	± 2.97

* Significantly different from mean of parous 4 ($p < .05$).
** Significantly different from mean of parous 3 ($p < .05$) and parous 4 ($p < .005$).

ing," "digestion," and "bowel movements." On the other hand, no significant differences were found with regard to concern about such items as "abnormalities of the navel," "buttocks," or "weight gain." The incidence of worries about "accidents" actually increased significantly with the birth of a second child, but returned to primiparous levels with the presence of four or more children in the family. Of particular interest is the fact that the incidence of concern about colic remained constant (approximately 11 per cent at all levels of parity up through seven); an increase in maternal parity did not alter the frequency of expression of concern about colic.

Variables within the infant such as sex, birth weight, or age in weeks as it ranged from five to ten were not found to be associated with significant differences in the mean number of total concerns indicated. However, several specific concerns did show significant fluctuations in association with sex differences in the infants; these will be discussed in a subsequent paper.

Variation in the geographical location of the respondents, with regard to region of the country or to "urban" or "rural" classification as herein employed, was not found to be associated with significant fluctuations in total number of concerns expressed.

DISCUSSION

Mail Questionnaire Surveys

Prior to interpretation of these results adequate consideration must be given to the certain recognized limitations of the mail questionnaire technique. In this particular instance, certain measures were undertaken to minimize their impact.

1. The written word cannot be assumed to have the same connotation to the reader that the spoken word has for the listener. Neither can it be assumed to have a universal connotation among the members of any one

group within the population. Therefore, where possible, potentially confusing or emotionally charged words or phrases were avoided, eliminated, or reduced to a minimum in the questionnaire.

2. The sequential listing of items for consideration is known to influence the testee in completing a questionnaire. In an attempt to minimize this source of error, those sources of concern noted most frequently during preliminary phases of constructing the questionnaire were spread throughout its final form.

3. Clarity of directions for accurate completion of a questionnaire must be included to minimize misinterpretation. Pretesting was, therefore, undertaken to confirm the adequacy of instructions.

4. A precise quantitation by mail questionnaire of the intensity of a specific emotion is obviously impossible. Nevertheless, the respondents were offered the opportunity to note one of two quantifying adjectives, "some" and "considerable." In analyzing the data for correlative relationships, however, the total number of concerns, regardless of indicated intensity, was utilized.

5. While desirable, precise differentiation of concerns based on "objective reality" as contrasted to those based on "subjective fantasy" is also obviously not feasible. Consequently, all concerns, regardless of apparent etiology, were included in the total used for correlative purposes.

6. The absence of a specific category in which to place her source of concern might lead a mother to (a) misplace the source in another category, (b) exclude the specific source altogether, or (c) fail to respond with a completed questionnaire. Consequently, a space for comments was included with the hope that none of these alternatives would be realized.

7. A questionnaire based on leading questions generally elicits a greater total number of answers than does one based on open-ended questions. By calling attention to specific items, it may actually entice the respondent to check an excess number of items. Such must be considered in interpreting the results of questionnaires; they reflect *relative* rather than *absolute* values.

8. Use of a selected group for sampling without an awareness of this by the tester invalidates any generalizations derived from resultant data. The group sampled was recognized as being selected on the basis of their all being mothers of young infants and their all having mailing addresses. Similarly, respondents automatically become still further selected in that they returned the questionnaire. No direct information is, therefore, available regarding the group that failed to respond. However, because of the consistency in incidence of concern between responses to the first and second mailings, further extrapolation from the data may possibly be warranted.

In spite of these inherent limitations, the mail questionnaire survey, by eliminating certain variables that pertain to the personal interview, can still provide meaningful data provided a high percentage of returns is obtained. The percentage of returns in both surveys ($60.5 + 47.5 = 53$ per cent) is considerably higher than that anticipated in a mail questionnaire survey which is largely nonselected. The primary reason, no doubt, is probably to be found in the high degree of interest possessed by potential respondents in the topic under survey. Admittedly, a still higher percentage of returns would be desirable, but the comparability of replies to first and second mailings suggests that, for the purposes proposed, the technique is satisfactory.

That the second survey, employing an identical questionnaire, duplicated the results of the first so closely evidences the reproducibility of findings. Also, the fact that for both surveys the distribution of returns compared closely with that of control data (except with regard to level of education) would suggest this technique constitutes an economical but unexploited method for gathering certain childrearing data from widespread sources during a brief span of time.

Mothers' Worries

As illustrated in Table 1, the relative frequency with which responding mothers noted various sources of concern is in general agreement with previous more limited observations. Carithers (3), in tabulating the concern among mothers in his practice, as these were expressed during the period of hospital confinement, found that "physiologic" outnumbered "anatomic" areas as sources of concern by a ratio of 4:1. In his study, worry about the infant's gastrointestinal tract and the skin was also most common. Montgomery (7), too, found the gastrointestinal tract to be the central focus of concern. When he contrasted the incidence of "chief complaints" among a group of mothers during the period from 1926 to 1930 with that from another group of mothers during the period of 1946 to 1949, he found that six leading complaints in the first group all related to gastrointestinal function and that five of these six categories continued to predominate as sources of concern in the latter group.

On the other hand, Escalona (4) has found that the baby's crying was the most frequent source of concern. But, as she has pointed out, this is often interpreted by the mother as representing hunger and hence is assumed by her to represent a gastrointestinal aberration. Escalona also notes the frequent finding of other concerns relating more directly to the gastrointestinal tract, such as the infant's vomiting, constipation, etc. She adds that, in this culture, mothers seem better able to relate these particular concerns to their physicians than concerns about the baby's breathing, bathtime, or startle reflex.

Blum (1), working in a well baby clinic that supervised care during the first year of life, also found that "hunger" and "excessive crying" were the most frequently expressed sources of concern. Again, both of these were interpreted by the mothers in her series as relating to gastrointestinal function or malfunction.

It would seem pertinent to contemplate the influence that medicine in general, and pediatrics in particular (with its origins so closely intertwined with the development of "safe infant nutrition"), has had or continues to have in possibly encouraging the mother to focus her own anxiety on the infant's gastrointestinal tract. Perhaps a reorientation emphasizing other factors in the infant or in his external environment would be advisable.

Influence of Maternal Factors

Among the maternal factors—age, level of education and parity—only parity was significantly associated with fluctuation in expression of concern. An increase in maternal parity was reflected in a reduction of concerns expressed. Blum has also reported that, on the basis of the total concerns they expressed, multiparous mothers tended to be less anxious or worried than primiparous mothers. Carithers, however, concluded multiparity was not associated with a decrease in total concerns as expressed by subjects in his series. In a differing approach to the influence of parity, Brody (2) separated a group of mothers into four tentative subgroups on the basis of extensive observational and psychological testing criteria. In subgroup A the mothers "were conspicuous for their ability to accommodate to the needs of their infants," while those in subgroup D "were conspicuously active but also erratic in their attentiveness, efficiency and sensitivity. They quite sedulously governed their infants' actions by stimulating, restricting or instructing them, apparently hardly aware of the possible effects of their behavior on the infants' condition . . ." (pp. 265-266). Of interest is the fact that among the seven mothers in subgroup A only one was primiparous, while among those in group D eight of eleven were primiparous, a statistically significant contrast. The present investigation would support her interpretation that with increase in parity, and consequently in experience, there is a decrease in expressed maternal anxiety. How this may relate to currently popular theories on the etiology of colic, the incidence of which remained constant at all levels of parity will be discussed in a subsequent paper.

Among the specific concerns, the observed initial increase in concern about accidents associated with an early increase in parity, and its subsequent return to primiparous levels as parity increased, would seem to reflect experiential learning on the part of the mother. The marked decrease in incidence of concern about bowel movements, sleeping, spoiling, and crying associated with increase in parity would also suggest that learning by experience can be successful in reducing a mother's anxiety. On the other hand, the relative stability of concern about the navel or the buttocks would suggest that, in spite of the mother's learning by experience, specific items continue to represent sources of concern to her. Such a finding makes more understandable the observation that, while the physician's brief explanation about one source of concern may calm an anxious mother, his detailed and repeated explanations about another may be totally without effect. It also suggests that, while anticipatory instruction or education about certain aspects of baby care or childrearing might be of considerable value, that about others would prove of little or no benefit.

While cautious interpretation is no doubt called for, other implications may be recognized in the results of this investigation. The absence as well as the presence of associations form the basis for many postulates. For example, the lack of association between fluctuation in the total number of concerns expressed by mothers and variables within their infants suggests that the focus of child health supervision might profitably include more of the mother. It would seem to confirm beliefs that she, rather than he, accounts for much of the contrast in emphasis of the infant's history. In such instances, therapeutic intervention aimed solely at the infant as the source of concern would be doomed to failure.

Recently Ross, in response to an editorial plea to pediatricians for "truly competent counselling in mental hygiene" (6), has emphasized the tenuousness of any claims for a scientific basis in such endeavors, pointing out that the available reference literature "is full of conjecture, speculation and personal impression" and derived from "relatively limited socio-economic groups" (8). When scrutinized, much currently popular childrearing advice —often given under the guise of child health supervision—fares little better, in terms of validity, than does preventive psychiatry. It, too, frequently stems from contradictory proclamations of intuitive dogma unsupported by data. It is suggested that selected use of the mail questionnaire survey—in addition to more widely accepted devices—be employed more extensively in attempts to gather objective data in this field. Such might preclude the obvious "reversal of field" in childrearing advice that has transpired in the past two generations from occurring in the future.

SUMMARY AND CONCLUSIONS

Two mail questionnaire surveys were conducted among nationwide samples of mothers of young infants to determine the incidence of their various concerns. The results were comparable. The mothers were found to be nonselected except, as was anticipated, with regard to their level of education.

Concerns about gastrointestinal dysfunction, skin disorders, and sleep disturbances were most prevalent. Fluctuations in incidence were analyzed for association with such variables as maternal age, education, and parity, infant's sex, and geographical location of the family. Only an increase in maternal parity was associated with a reduction in incidence of concerns.

It is suggested that the technique of mail questionnaire survey offers an effective and economical method for collecting objective data. Wider usage should be considered in approaching problems relating to comprehensive child health supervision.

APPENDIX

DUPLICATE OF STUDY QUESTIONNAIRE

Baby's date of birth ____________________ Today's date ________________

Baby's weight at birth __________________ Mother's age ________________

Baby is a boy ☐ girl ☐ How many *other* children do you have? ____

The following list is based on doctors' reports of the many questions or worries that mothers sometimes have about their new babies. Which of these, if any, have *worried* you about your baby?

	No Worry (check)	Some Worry (check)	Considerable Worry (check)	Please Describe
Stomach (too large, small, hard, soft, swollen, etc.)	☐	☐	☐	________________ ________________
Breathing (uneven, hiccoughs, gags, chokes, gasps, grunts, etc.)	☐	☐	☐	________________ ________________
Hair (too much, too little, falling out, etc.)	☐	☐	☐	________________ ________________
Eyes (puffy, red, crossed, color, etc.)	☐	☐	☐	________________ ________________
Ears (shape, size, color, etc.)	☐	☐	☐	________________ ________________
Nose (size, shape, running, etc.)	☐	☐	☐	________________ ________________
Mouth-Lips (size, shape, color, sore, swallowing, thumb-sucking, etc.)	☐	☐	☐	________________ ________________
Sleeping (too much, not enough, not regular, restless, etc.)	☐	☐	☐	________________ ________________
Accidents (while sleeping, eating, playing, bathing, etc.)	☐	☐	☐	________________ ________________
Head (size, shape, soft spot, etc.)	☐	☐	☐	________________ ________________
Weight (not gaining, too fat, too thin, etc.)	☐	☐	☐	________________ ________________
Crying (too much, too little, strong, weak, turns color, etc.)	☐	☐	☐	________________ ________________

(*Study Questionnaire continued on next page*)

DUPLICATE OF STUDY QUESTIONNAIRE (continued from previous page)

NAVEL (swollen, too large, small, bleeding, odor, etc.)	☐	☐	☐	________
BUTTOCKS (diaper rash, sore, color, etc.)	☐	☐	☐	________
SKIN (oily, dry, rash, scratches, etc.)	☐	☐	☐	________
BOWEL MOVEMENTS (odor, color, too often, too loose, hard, etc.)	☐	☐	☐	________
URINE (odor, color, too often, too little, etc.)	☐	☐	☐	________
EATING (not enough, too much, not regular, hungry, disagrees, etc.)	☐	☐	☐	________
DIGESTION (spitting, burping, gas, vomiting, colic, etc.)	☐	☐	☐	________
LEGS-FEET (too thin, too heavy, not straight, etc.)	☐	☐	☐	________
ARMS-HANDS (too thin, too heavy, not straight, etc.)	☐	☐	☐	________
OTHER (spoiling baby, food preparation, bathing, clothing, diapering, etc.)	☐	☐	☐	________

Have you chosen a doctor to care for your baby yet? Yes ☐ No ☐

If yes, has the doctor examined your baby in—
(Please check) a. the hospital ☐ b. the office ☐
c. your home ☐ d. baby not examined by doctor yet ☐

Did the doctor mention anything about your baby that would need special care or attention? (Please explain) ________________________________
____
____

About yourself, how many years were you in— Grade School ____ yrs.; High School ____ yrs.; College ____ yrs.; Postgraduate ____ yrs.

Comments ________________________________
____
____
____
____
____
____
____

REFERENCES

1. BLUM, L. H. Some psychological and educational aspects of pediatric practice: A study of well-baby clinics. *Genet. Psychol. Monogr.*, 1950, 41, 3-97.
2. BRODY, S. *Patterns of mothering.* International Universities Press, 1956.
3. CARITHERS, H. A. Psychologic aspects of pediatrics: Mother-pediatrician relationship in the neonatal period. *J. Pediat.*, 1951, 38, 654-660.
4. ESCALONA, S. Problems of infancy and childhood. In M. J. E. Senn (Ed.), *Transactions of third conference.* Josiah Macy, Jr. Foundation, 1949. Pp. 30-95.
5. MAY, C. D. Editorial: Can the new pediatrics be practiced? *Pediatrics*, 1959, 23, 253.
6. MAY, C. D. Editorial: Mental hygiene in the new pediatrics. *Pediatrics*, 1959, 24, 355.
7. MONTGOMERY, Z. D. Problems of infancy and childhood. In M. J. E. Senn (Ed.), *Transactions of third conference.* Josiah Macy, Jr. Foundation, 1949. Pp. 11-29.
8. ROSS, L. S. Letters to the editor. *Pediatrics*, 1960, 25, 174-175.
9. SHAFFER, T. E. *Hospital acquired staphylococcal disease; example of questionnaire survey of discharged infants.* Proc. Natl Conf. Dept. Hlth, Educ. Welf., 1958.
10. SNEGIREFF, L. S., & LOMBARD, O. M. Smoking habits of Massachussetts physicians: Five-year follow-up study (1954-1959). *New England J. Med.*, 1959, 261, 603-604.
11. Annual Report: Vital statistics of the U. S., 1954, Vol. I. Dept. Hlth, Educ. Welf., Public Hlth Service, Natl Office Vital Statistics, 1955.
12. Annual Survey for 1956, Part I, Monthly vital statistics report. Dept. Hlth, Educ. Welf., Public Hlth Service, Natl Office Vital Statistics, 1957.
13. Current population reports. Bureau of Census, Washington, D.C., March, 1957.

POSITIVE REINFORCEMENT AND BEHAVIORAL DEFICITS OF AUTISTIC CHILDREN

C. B. FERSTER*

Indiana University Medical Center

Infantile autism, first described by Kanner (6), is a very severe psychotic disturbance occurring in children as young as 2 years. At least outwardly, this childhood schizophrenia is a model of adult schizophrenia. Speech and control by the social environment are limited or absent; tantrums and atavistic behaviors are frequent and of high intensity; and most activities are of a simple sort, such as smearing, lying about, rubbing a surface, playing with a finger, and so forth. Infantile autism is a relatively rare form of schizophrenia and is not important from an epidemiological point of view. The analysis of the autistic child may be of theoretical use, however, since his psychosis may be a prototype of the adult's; but the causal factors could not be so complicated, because of the briefer environmental history. In this paper, I should like to analyze how the basic variables determining the child's behavior might operate to produce the particular kinds of behavioral deficits seen in the autistic child. To analyze the autistic child's behavioral deficits, I shall proceed from the general principles of behavior, derived from a variety of species, which describe the kinds of factors that alter the frequency of any arbitrary act (3, 10). The general principles of behavior applied to the specific situations presumably present during the child's developmental period will lead to hypotheses as to specific factors in the autistic child's home life which could produce the severe changes in frequency as well as in the form of his behavior. As an example, consider the effect of intermittent reinforcement, many of the properties of which are comparatively well known from animal experiments. To find how intermit-

---

* Institute of Psychiatric Research, Indiana University Medical Center, 1100 West Michigan Street, Indianapolis 7.

Child Develpm., 1961, 32, 437-456.

tent reinforcement of the autistic child's behavior might produce deficits, we would first determine, in the general case, what specific orders of magnitude and kinds of schedules produce weakened behavioral repertoires. The factors in the child's home life could be examined to determine estimates of what kind of circumstances could conceivably cause schedules of reinforcement capable of the required attenuation of the child's behavior. The analysis will emphasize the child's performance as it is changed by, and affected in, social and nonsocial environment. As in most problems of human behavior, the major datum is the frequency of occurrence of the child's behavior. Although the account of the autistic child's development and performance is not derived by manipulative experiments, it may still be useful to the extent that all of the terms of the analysis refer to potentially manipulable conditions in the child's environment and directly measurable aspects of his performance. Such an analysis is even more useful if the performances and their effects on the environment were described in the same general terms used in systematic accounts of behavior of experimental psychology.

Some of our knowledge of the autistic child's repertoire must necessarily come from anecdotal accounts of the child's performance through direct observation. Although such data are not so useful as data from controlled experiments, they can be relatively objective if these performances are directly observable and potentially manipulable. A limited amount of experimental knowledge of the dynamics of the autistic child's repertoire is available through a program of experiments in which the autistic child has developed a new repertoire under the control of an experimental environment (5). These experiments help reveal the range and dynamics of the autistic child's current and potential repertoires. In general, the autistic child's behavior will be analyzed by the functional consequences of the child's behavior rather than the specific form. The major attempt will be to determine what specific effects the autistic child's performance has on that environment and how the specific effects maintain the performance.

SPECIFICATION OF THE AUTISTIC CHILD'S PERFORMANCE

We must first describe the current repertoire of the autistic child before we can describe possible environmental conditions that might produce gross behavioral deficits. A topographic description of the individual items of the autistic child's repertoire would not, in general, distinguish it from the repertoires of a large number of functioning and nonhospitalized children, except perhaps in the degree of loss of verbal behavior. The autistic child's behavior becomes unique only when the relative frequency of occurrence of all the performances in the child's repertoire is considered. In general, the usual diagnostic categories do not adequately characterize the children in the terms of a functional analysis of behavior. Hospitalization of a child usually depends upon whether the parent can keep the child in the home, rather than a functional description of the role of the parental environment in sustaining or weakening the child's performance.

Range of Performances

Although the autistic child may have a narrower range of performances than the normal child, the major difference between them is in the relative frequencies of the various kinds of performances. The autistic child does many things of a simple sort—riding a bicycle, climbing, walking, tugging on someone's sleeve, running, etc. Nevertheless, the autistic child spends large amounts of time sitting or standing quietly. Performances which have only simple and slight effects on the child's environment occur frequently and make up a large percentage of the entire repertoire, for example, chewing on a rubber balloon, rubbing a piece of gum back and forth on the floor, flipping a shoelace, or turning the left hand with the right. Almost all of the characteristic performances of the autistic child may be observed in nonhospitalized children, but the main difference lies in the relative importance of each of these performances in terms of the total repertoire. Conversely, isolated instances of quite "normal" performances may be seen in the autistic child. Again, the relative frequency of the performances defines the autistic child.

Social Control over the Child's Performance

The major performance deficits of the autistic child are in the degree of social control: The kinds of performances which have their major effects through the mediation of other individuals.

The main avenue of social control in a normal repertoire is usually through speech, a kind of performance that is unique because it produces the consequences maintaining it through the mediation of a second person (12). Autistic children almost always have an inadequately developed speech repertoire, varying from mutism to a repertoire of a few words. Even when large numbers of words are emitted, the speech is not normal in the sense that it is not maintained by its effect on a social environment. When normal speech is present, it usually is in the form of a *mand* (12). This is a simple verbal response which is maintained because of its direct reinforcement, e.g., "Candy!" "Let me out." The main variable is usually the level of deprivation of the speaker. It lacks the sensitive interchange between the speaker and listener characteristic of much human verbal behavior, as for example, the *tact* (*see* below). The reinforcement of the mand largely benefits only the speaker. In the case of the autistic child, it frequently affects the listener (parent), who escapes from the aversive stimulus by presenting a reinforcing stimulus relevant to the child's mand. At suppertime, the child stands at the door screaming loudly and kicking the door because the ward attendants in the past have taken the child to supper when this situation became aversive enough. Sometimes, the form of the mand is nonvocal, although still verbal, as when the mand involves tugging at a sleeve, pushing, or jostling. The dynamic conditions which could distort the form of a mand into forms most aversive to a listener will be described below. In contrast to the mand, the tact (12) is almost completely absent. This form of verbal behavior benefits the listener rather than the speaker and is not usually relevant to the current deprivations of the speaker. This is the form of verbal behavior by which the child describes his environment, as, for example, "This is a chair"; "The mailman is coming." This latter kind of verbal control is generally absent or weak, as with other kinds of verbal behavior except an occasional mand.

Atavisms

Tantrums, self-destructive behavior, and performances generally aversive to an adult audience are relatively frequent in the autistic child's repertoire. Most autistic mands depend on an aversive effect of the listener for their reinforcement. To the extent that social behavior is present at all, its major mode is through the production of stimuli or situations which are aversive enough so that the relevant audience will escape or avoid the aversive stimulus (often with a reinforcer). For example, on the occasion of candy in the immediate vicinity, the child screams, flails about on the floor, perhaps striking his head, until he is given some candy. There is evidence that much of the atavistic performance of the autistic child is operant, that is, controlled by its consequence in the social environment. The operant nature of the autistic child's atavisms is borne out by experiments where a child was locked in an experimental space daily for over a year. There was no social intervention, and the experimental session was usually prolonged if a tantrum was underway. Under these conditions, the frequency of tantrums and atavisms declined continuously in the experimental room until they all but disappeared. Severe tantrums and attempts at self-destruction still occurred when sudden changes in the conditions of the experiment produced a sudden change in the direction of nonreinforcement of the child's performances. Severe changes in the reinforcement contingencies of the experiment produced a much larger reaction in the autistic than in the normal child. Consequently, we learned to change experimental conditions very slowly, so that the frequency of reinforcement remained high at each stage of the experiment. Much of the atavistic behavior of the autistic child is maintained because of its effect on the listener.

Reinforcing Stimuli

The reinforcers maintaining the autistic child's performance are difficult to determine without explicit experimentation. Small changes in the physical environment as, for example, direct stimulation of the mouth, splashing water, smearing a sticky substance on the floor, breaking a toy, or repeated tactile sensations, appear to sustain the largest part of the autistic child's repertoire. Nevertheless, these may be weak reinforcing stimuli which appear to be strong, because the response produces its reinforcement continuously and because alternative modes of responding are also maintained by weak reinforcers. The durability and effectiveness of a reinforcer can usually be determined best by reinforcing the behavior intermittently or by providing a strong alternative which could interfere with the behavior in question. In the controlled experiments with autistic children, most of the consequences we supplied to sustain the children's performance, such as color wheels, moving pictures, music, and so forth, were very weak reinforcers compared with food or candy. Food generally appeared to be an effective reinforcer, and most of the performances associated with going to the dining room and eating are frequently intact. In contrast, the normal children could sustain very large amounts of behavior through the nonfood reinforcements. It is difficult to guess the potential effectiveness of new reinforcers, because the estimate depends upon some performance being maintained by that reinforcer.

In the everyday activities of the autistic children, little behavior was sustained by conditioned or delayed reinforcers. But, in a controlled experimental situation, such activities could be sustained by explicit training. For example, (a) The sound of the candy dispenser preceding the delivery of candy served as a conditioned reinforcer. The fine-grain effects of the schedules of reinforcement show this. The difference in performance produced by two different schedules of reinforcement could have occurred only if the effective reinforcer were the sound of the magazine rather than the delivery of a coin. The actual receipt of the coin or food is much too delayed to produce the differences in performances under the two schedules without the conditioned reinforcer coming instantly after a response. (b) With further training, the delivery of a coin (conditioned reinforcer) sustained the child's performances. The coin, in turn, could be used to operate the food or nonfood devices in the experimental room. (c) Still later, coins sustained the child's performance even though they had to be held for a period of time before they could be cashed in. The child worked until he accumulated five coins, then he deposited them in the reinforcing devices. (d) Even longer delays of reinforcement were arranged by sustaining behavior in the experimental room with a conditioned reinforcer as, for example, a towel or a life jacket which could be used later in the swimming pool or in water play after the experimental session terminated. The experimental development of these performances shows that, even though the usual autistic repertoire is generally deficient in performances sustained by conditioned reinforcement and with delay in reinforcement, the children are potentially capable of developing this kind of control.

Little of the autistic child's behavior is likely to be maintained by generalized reinforcement, that is, reinforcement which is effective in the absence of any specific deprivation. A smile or parental approval are examples. The coins delivered as reinforcements in the experimental room are potentially generalized reinforcers, since they make possible several performances under the control of many different deprivations. However, we do not know whether the coin has actually acquired the properties of a generalized reinforcer.

Stimulus Control of Behavior

It is very difficult to determine the stimulus and perceptual repertoire of autistic children. When a child responds to a complex situation, it is not usually clear what aspect of the situation is controlling the child's behavior. In most cases, it is difficult to determine to what extent these children can respond to speech discriminatively, since the situations are usually complex and many stimuli may provide the basis for the simple performances. Similarly with visual repertoires. Controlled experiments showed unequivocally that behavior can come under the control of simple stimuli when differential effects of the performances were correlated with the different stimuli. When a coin was deposited in a lighted coin slot, it operated the reinforcing device. Coins deposited in unlighted slots were wasted. The children soon stopped putting coins in the unlighted slots. The previously developed stimulus control broke down completely when these stimuli were placed in a more complicated context, however. A new vending machine was installed with eight columns, eight coin lights, and eight coin slots, so the child could choose a preferred kind of candy. The slight increase in complexity disrupted the control by the coin light, and it took several months and many experimental procedures before the stimulus control was reestablished. A better designed procedure, in view of the minimal perceptual repertoire of these children, would have been a gradual program by which variations in the specific dimensions of the coin slot and coin light were changed while the reinforcement contingency was held constant in respect to the essential property.

In summary, the repertoire of the autistic child is an impoverished one. Little is known about the perceptual repertoire, but the available evidence suggests that it is minimal. The absolute amount of activity is low, but this deficit is even more profound if the specific items of activity are evaluated in terms of whether they are maintained by significant effects on a social or even nonsocial environment. Most of the child's performances are of a simple sort, such as rubbing a spot of gum back and forth, softening and twisting a crayon, pacing, or flipping a shoelace. Those performances in the child's repertoire having social effects frequently do so because of their effects on the listener as aversive stimuli. Atavisms and tantrums are frequent.

THE EMERGENCE OF PERFORMANCE DEFICITS DURING THE EARLY DEVELOPMENT OF THE AUTISTIC CHILD

Having characterized the autistic child's repertoire, the next step is to determine the kinds of circumstances in the early life of these children which could bring about the behavioral deficits. The general plan is to state how the major behavioral processes and classes of variables can drastically reduce the frequency of occurrence of the various behaviors in the repertoire of any organism. Then, the parental environment will be examined to determine circumstances under which the actual contingencies applied by the parental environment to the child's behavior could weaken the child's performance similarly. The datum is the frequency of occurrence of all of the acts in the child's repertoire, and the independent variables are the consequences of these acts on the child's environment, particularly the parental environment. All of the terms in such a functional analysis are actually or potentially directly observable and manipulable. In general, the performances in the child's repertoire will be simultaneously a function of many factors, each contributing to changes in the frequency of the relevant performances. It is important, therefore, to consider relative changes in frequency rather than simple presence or absence of various performances. The datum is the frequency of occurrence of the behavior. In the same vein, singly identifiable factors may be interrelated and functioning simultaneously.

The major paradigm for describing the behavior of an organism is to specify the consequences of the act (reinforcement) which are responsible for its frequency. In this sense, the major cause of an instance of behavior is the immediate effect on the environment (reinforcement). The continued emission of the verbal response "Toast" depends on its effect on the parent in producing toast. Every known behavioral process influencing the frequency of a positively reinforced performance is relevant to the problem of defining conditions under which we may produce a behavioral deficit. Given the variables which maintain it, a performance may be weakened by their absence or by changing the order of magnitude. It is perhaps surprising to discover that large behavioral deficits are plausible without any major appeal to punishment or suppression of behavior by aversive stimuli.

Intermittent Reinforcement and Extinction

Intermittent reinforcement and extinction are the major techniques for removing or weakening behavior in a repertoire. The most fundamental way to eliminate a kind of behavior from an organism's repertoire is to discontinue the effect the behavior has on the environment (extinction). A performance may also be weakened if its maintaining effect on the environment occurs intermittently (intermittent reinforcement).[1] Behaviors occurring because of their effects on the parent are especially likely to be weakened by intermittent reinforcement and extinction, because the parental reinforcements are a function of other variables and behavioral processes usually not directly under the control of the child. The reinforcement of the verbal response, "Give me the book," may go unreinforced because of many factors which determine the behavior of the listener. He may be preoccupied, listening to someone else, disinclined to reinforce, momentarily inattentive, etc. In contrast, the physical environment reinforces continuously and reliably. Reaching for a book is usually followed by the tactile stimulation from the book. Verbal behavior, particularly, depends entirely for its development and maintenance on reinforcements supplied by an audience (usually a parent). Because of the possibility of prolonged extinction and infrequent, intermittent reinforcement, speech and social behavior are the most vulnerable aspects of the child's repertoire. The young child is particularly vulnerable to the extinction and intermittent reinforcement occurring in social reinforcement because only the parental environment mediates nearly all of the major reinforcers relevant to his repertoire. Large parts of the child's repertoire are reinforced by first affecting a parent who in turn produces the reinforcer for the child. The 2-year-old child who asks for a cookie from a parent and gets no response usually has no alternative audience who will reinforce this vocal behavior. The result will either be the extinction of the child's verbal behavior or the reinforcement of nonvocal verbal forms when the child produces a cookie by a tantrum from which the parent escapes by giving the cookie.

---

[1] The reader may suggest at this point an apparent contradiction with the fact that extinction after intermittent reinforcement is more prolonged than after continuous reinforcement. This aspect of intermittently reinforced behavior's durability is not a general proposition, however, and does not hold for behavior which is still being maintained. Behavior reinforced intermittently will, in general, be emitted less frequently and be more easily weakened by emotional factors, changes in deprivation, punishment, and physiological disturbances than continuously reinforced behavior.

Factors in the Parental Repertoire Affecting the Frequency of Reinforcement of the Child's Performances

To find the conditions under which the child's repertoire will be weakened, therefore, we must look for conditions influencing the parents' behavior, which will alter the parental performances, in turn providing reinforcement of the child's performances. These might be:

1. The general disruption of the parental repertoire. Any severe disruption of the parental repertoire will severely affect the frequency with which the parent reinforces the behavior of the child. Consider, for example, the depressed parent whose general level of behavior is very low. One consequence of this low level of behaving will be a lessened frequency of reacting to the child. Therefore, many items in the child's repertoire will be less frequently reinforced in the depressed than the normal parent. The verbal responses, "May I have some bread" or "I want to go outside," might go unreinforced or be emitted many times without reinforcement. Various kinds of somatic disturbances, such as alcoholic "hangover," drug addiction, severe headache, somatic diseases, etc., could also produce large changes in the over-all reactivity of the parent to a child. To the extent that the child's performances occur because of their effect on the parent, the severely weakened parental repertoire may correspondingly weaken the child's behavior. If the parental extinction of the child's behavior is systematic and periodic, much of a child's behavior could be eliminated.

2. Prepotency of other performances. Whether or not a parent reinforces a child's performance also depends upon the alternative repertoire available to the parent. For example, the parent who is absorbed in various kinds of activities such as housecleaning, a home business, social activities and clubs, active telephoning, and so forth, may at various times allow many usually reinforced performances to go unreinforced. In general, the likelihood of omitting reinforcement would depend upon the strength of the prepotent repertoire. As an example of a prepotent repertoire, the housewife absorbed in a telephone conversation will not be inclined to answer a child or comply with a request. Housecleaning might be another repertoire controlling some parents' behavior so strongly that it is prepotent over behavior in respect to the child. In both cases, the essential result is the nonreinforcement of the child's behavior in competition with the prepotent parental repertoire. Mothers of autistic children often appear to have strong repertoires prepotent over the child. This may be at least a partial reason why mothers of autistic children are so often well-educated, verbal, and at least superficially adequate people.

3. A third factor producing intermittent reinforcement of the child's behavior is related to the first two factors listed above. If the parent finds other reinforcers outside of the home more rewarding than dealing with the child, the child becomes an occasion on which the significant elements of the parental repertoire cannot be reinforced. A parent changing diapers, or otherwise taking care of a child, cannot telephone a friend, be out socializing, be on a job, or doing whatever the autistic mother finds rewarding. The child acquires the properties of a conditioned aversive stimulus because it is an occasion which is incompatible with the parents' normal repertoire. This is of course the major method of aversive control in human behavior—the discontinuation of positive reinforcement. Another basis for establishing the child as a conditioned aversive stimulus to the parent is the emergence of atavisms and a large degree of aversive control of the parent by the child. To the extent that the parent is reinforced by escaping from the child because of his conditioned aversive properties, the frequency of the parental reinforcement of the child's behavior is further reduced.

The development of the atavistic behavior in the child by the parent is necessarily a very gradual program in which the beginning steps involve small magnitudes of behavior such as whining, whimpering, and crying. As the parent adapts to these or becomes indifferent to them because of the prepotence of other kinds of activity, then progressively larger orders of magnitude become reinforced. The large-magnitude tantrum may be approximated or "shaped" by gradual differential reinforcement. The parents of one autistic child, for example, at one period took turns all night standing in the child's room because one step out of the room would immediately produce a severe tantrum in the child. When the child functions as a conditioned aversive stimulus for the parent, the parent is less likely to reinforce the child's behavior positively. This lack of positive reinforcement, in turn, emphasizes the atavistic responses on the child's part as the major mode of affecting the parent.

The usual limiting factor in preventing excessive development of tantrums is the emergence of self-control on the part of the parent in escaping from the aversive control by the child rather than reinforcing it. Here, again, the repertoire of the parent is relevant. The development of self-control requires a highly developed repertoire which depends for its development on the ultimate aversive consequences of the child's control of the parent. The child's control becomes more aversive to the parent if it interrupts strong repertoires. Specifically, a parent engrossed in a conversation will find a child's interruption more aversive than a parent who is simply resting. If, in fact, there is no strong behavior in the parent, then the child's control is not likely to be aversive, and there is no basis for developing self-control.

All three of the above factors—over-all disturbances in the parental repertoire, prepotent activities, and escape from the child because of his aversiveness—reduce the amount of parental reinforcement of the child's performances. The over-all effect of the nonreinforcement on the repertoire of the child will depend upon the length of time and number of items of the child's repertoire that go unreinforced, as well as the existence of other possible social environments that can alternatively maintain the child's behavior (*see* below).

Differential Reinforcement of Atavistic Forms of Behavior by the Parent

The schedule by which the parent reacts to the child is also relevant to the development of atavistic behavior. Initially, a tantrum may be an unconditioned consequence of parental control as, for example, sudden nonreinforcement or punishment. Eventually, however, the child's tantrums may come to be maintained by their effect on the parental environment, because they present an aversive situation that can be terminated if the parent supplies some reinforcer to the child. The reinforcer presented by the parent to escape from the aversive consequences of the tantrum also increases the subsequent frequency of atavistic responses.

The effect on the parent of the given form and intensity of tantrums will vary from time to time, depending on the conditions maintaining the parents' behavior. This variation in sensitivity of the parent to aversive control by the child results in a variable-ratio schedule of reinforcement of the child's tantrum by the parent—a schedule of reinforcement potentially capable of maximizing the disposition to engage in tantrums. This is the schedule of reinforcement that produces the high frequencies of performances as in gambling (10). The sensitivity of the parent to aversive control by the child will depend on the general condition of the parental repertoire as discussed above. The same factors in the parental repertoire that tend to produce nonreinforcement of the child's behavior—general disruption of the parent or other behaviors prepotent over the child—correspondingly produce reinforcement of large-order-of-magnitude tantrums. The parent whose total repertoire is severely enough disrupted to interfere with the normal reinforcement of the child's behavior will also react only to tantrums that are of large order of magnitude of aversiveness. A range of sensitivity of the parent to aversive control by the child produces ideal conditions for progressively increasing the intensity or frequency of tantrums. A high sensitivity to aversive control guarantees that some tantrums will be reinforced at least periodically. A low sensitivity differentially reinforces tantrums of large orders of magnitude. At one extreme, the parent may be hypersensitive to the child and, at other times, so depressed that only physical violence will produce a reaction. The schedule by which the parent's behavior terminates the tantrum is a second factor which will increase the range of reactivity of the parent. As more behavior is required of the parent to terminate the tantrum, the parent's inclination to do so will fall. When the parent is less inclined to reinforce a given intensity of tantrum, any variation in tantrum intensity is tantamount to differential reinforcement of extreme forms, if the parent now reacts to the larger-order-of-magnitude tantrum.

How much the parent differentially reinforces tantrums in the child depends, in part, upon the child's other positively reinforced repertoires. When, for example, a child's performance suddenly goes unreinforced, as when a parent may refuse a request, the likelihood and severity of a tantrum will in part depend on the parent's ability to "distract" the child. This, in turn, depends upon whether alternative modes of behavior are in fact available to the child. When conditions are present for the progressive reinforcement of more and more severe tantrums, the process is potentially non-self-limiting. Autocatalysis is likely to occur, particularly if the parent has little disposition to reinforce the general items in the child's repertoire for reasons other than terminating the aversive demands of the child.

Nonsocial Reinforcers

Some of the child's behavior is maintained by his direct effect on the physical environment without the intervention of other individuals. In general, very small effects on the environment will sustain performances with which the parent usually has little reason to interfere. For example, the child plays with his own shoelace, moves his fingers in his own visual field, emits minimal nonverbal, vocal responses, and so forth. Larger effects on the physical environment as, for example, moving objects about the house, speaking to the parent, playing with toys, touching and handling usual household objects, are more likely to enter upon the parental repertoire and so may produce a response whose effect is to discontinue the behavior or interfere with its reinforcement. The punishment aspect of the parental interference with the child's activities will be dealt with separately below. The relative possibility of parental interference and nonreinforcement of the hierarchy of performances may account for the large part of the autistic child's repertoire, which consists of behaviors having small, limited effects on the physical environment. Occasionally, even behaviors that are maintained by the most simple effects on the environment are extinguished or punished when they occur in the presence of a parent. For example, the father of one autistic child reports that the child reached for a chandelier while he was holding him. The father instantly dropped the child, with a reaction of considerable disapproval because "You should pay attention to me when you're with me." Aside from the secondary effect on the child, the immediate result of the incident is the nonreinforcement of the child reaching for a common physical object.

The existence of "nonverbal" vocal behavior in some autistic children may be related to forms of vocal behavior with which the parent will or will not interfere. Vocal behavior maintained by its effect on a parent (verbal) is susceptible to weakening by parental extinction. A parent interferes less easily with vocal behavior maintained by its direct effect (nonverbal) comparable with making noise by rubbing a stick over a rough surface. Further, such nonverbal vocal responses can emerge readily at any stage of the child's life, unlike verbal behavior, because it does not depend on a generalized reinforcement.

Failure to Develop Conditioned and Generalized Reinforcers

The normal repertoire of the child consists almost entirely of sequences of behavior that are maintained, in a chain or sequence, by conditioned and generalized reinforcers (10). An example of a chain of responses would be the behavior of the child moving a chair across the room and using it to climb to a table top to reach a key which in turn opens a cupboard containing candy. This complicated sequence of behavior is linked together by critical stimuli which have the dual function of sustaining the behavior they follow (conditioned reinforcement) and setting the occasion for the subsequent response. The chair in the above example is an occasion on which climbing onto it will bring the child into a position where reaching for food on the table top will be reinforced by obtaining food. Once this behavior is established, the chair in position in front of the table may now be a reinforcer, and any of the child's behavior which results in moving the chair into position will be reinforced because of the subsequent role of the chair in the later chain of behaviors. A minimal amount of behavior is necessary before a chain of responses can develop. The development of the control by the various stimuli in the chain, both as discriminative stimuli setting the occasion for the reinforcement of behavior and as reinforcers, depends upon a high level of activity, so that the responses will occur and come under the control of the stimuli. This is even more true for the development of the generalized reinforcer. When the child has moved enough objects about the house and achieved a variety of effects on his environment relative to a range of deprivations and reinforcers, simply manipulating the physical environment may become a reinforcer without reference to a specific level of deprivation. This, of course, is the uniquely human reinforcer that makes possible much of verbal behavior, education in general, and self-control. Again, large amounts of behavior—many chains of behavior with many different kinds of conditioned reinforcers—are a necessary condition for the emergence of a generalized reinforcer. To the extent that the child's repertoire becomes weakened by intermittent reinforcement and extinction, as mentioned above, and punishment and aversive control (*see* below), the possibility of the development of generalized reinforcers, and hence more complex behavior, becomes less and less likely. Parental "attention" is probably one of the most important generalized reinforcers normally maintaining the child's behavior. Parental attention is an occasion upon which the child's performances may have an important effect on the parent. Inattention is an occasion on which the child's responses are likely to have little effect. Hence, the parents' performances in smiling, saying, "Right," "Good boy," or "Thank you," all come to function as conditioned reinforcers. Their emergence as generalized reinforcers again depends upon the existence of a large behavioral repertoire. A large number of chains of responses will produce important positive effects when the parent smiles or says, "Good boy." Lower frequencies of reinforcement follow for these same activities when the parent is frowning or says, "Bad boy."

Any large reduction in the child's over-all performance will interfere with the initial development of conditioned reinforcers or their continued effectiveness. The control by the environment over the child's behavior depends first upon the emission of the behavior. This follows from the manner in which the environment comes to control the child's performance: the successful execution of an act on one occasion, coupled with the unsuccessful act in its absence. Until a child climbs on chairs, as in the previous example, a chair has little chance of becoming a discriminative stimulus. Without the development of stimulus control, conditioned reinforcers cannot develop. The reinforcing effect of the chair in the above example depends upon its being the occasion on which further performances may be reinforced. In this way, a low general level of behavior may impede the enlargement of the child's repertoire because it does not allow stimulus control and in turn prevents reinforcement of new behavior. A limited development of simple conditioned reinforcers in turn prevents the development of a generalized reinforcer. Parental responses, such as smiling, "Good," or "Right," can have little effect on the child if there is not a history by which many different forms of the child's performance have produced various reinforcers on these occasions. Without parental generalized reinforcement, educational processes and positive parental control are all but impossible. This control is normally carried out by the use of praise and parental attention, coupled with mild forms of threats of discontinuing the reinforcers. Even after a generalized reinforcer has acquired its function, its continued effectiveness depends on the various stimuli continuing to stand in a significant relation to the child's performance. The actual form of the parents' generalized reinforcer is not nearly as important as the parents' subsequent reinforcement practices with the child. The reinforcing effects of the smile derive from the reinforcing practices associated with it. A smile usually functions as a generalized reinforcer in most people because a smiling person is more likely to reinforce. The correlation between smiling and reinforcement is by no means inevitable, however. Some individuals may be more disposed to punish than reinforce when smiling in some situations. In a similar vein, if the child has no behavior in his repertoire that will be more likely to be reinforced on the occasion of a parental smile, it matters little what the parent's reinforcing practices are when smiling as against when frowning.

STIMULUS CONTROL

The specific occasions on which a child's performances have their characteristic effects on the environment will subsequently determine whether the child acts. In the absence of the characteristic circumstances under which the behavior is normally reinforced, the child will be less disposed to act in proportion to the degree of similarity with the original situation. Changing a stimulus to one which has not been correlated with reinforcement is another way of weakening a repertoire. New stimuli also elicit emotional responses and general autonomic effects that may interfere with established performances. Here, simply repeated exposure to the stimuli may produce adaptation to the stimuli and eliminate their emotional effects. Ordinarily, the infants' performances are under the control of a limited range of stimuli, usually one or two parents in a limited part of a specific home environment. The discriminative repertoire broadens as the child grows older and other individuals come to be occasions on which his performances have significant effects. The parental environment of the very young child narrows the control of the child's performance to a limited range of stimuli, largely because the parent mediates almost all of the important events affecting the child. A major factor which brings the child's behavior more narrowly under the control of the parent is the nonreinforcement of much of the child's behavior in the absence of the parent. The close control of the child's behavior by the parent weakens the child's repertoire in the absence of a parent much more when there has been explicit differential reinforcement than when there has been simply a limited reinforcing environment.

Sudden shifts in the child's environment may or may not produce major performance deficits. At one extreme, a sudden shift of the stimuli in the child's controlling environment will have little influence if the child already has been reinforced on the occasion of a wide range of circumstances and individuals. At another extreme, a repertoire can be eliminated almost completely if the child has had a history in which major kinds of performances have gone unreinforced except on the occasion of a single person in a specific environment. The sudden shifts in the situations and persons controlling the child's behavior may occur under a variety of circumstances, such as a sudden change in a constant companion, death of a parent, or a sudden shift in the physical environment. A sudden shift in the environment of one of the subjects reported in the previously mentioned experiment could conceivably have been the major factor in her autistic development. Many of the activities of the child's mother were prepotent over dealing with the child, and she solved the problem by hiring a teenage baby sitter as a constant companion and nursemaid. After a year, the baby sitter left, suddenly and abruptly, leaving the child with the mother. Within four months, the child began to behave less in general, lost speech, and showed increasing frequency of atavisms. The child's repertoire possibly was under such close control of the baby sitter that the very sudden change to the mother created an environment which in the past had been correlated with nonreinforcement. If the child's behavior were under very narrow control by the baby sitter, because of the nonreinforcement on all other occasions, a sudden shift, as in the loss of the baby sitter, could produce a dramatic deficit in the child's repertoire.

Disruptive Effect of Sudden Stimulus Changes and the Amount, Durability, and Range of Behavior

A novel reinforcing environment will not sustain a child's performance unless the repertoire contains behavior of a sufficient range and durability. The new environment weakens the performance because it nearly always requires slightly different forms of behavior. For example, a new person entering a child's home is not so likely to respond successfully to the incompletely developed verbal behavior of a child as the parent. The possibility of the child's affecting the stranger will depend upon his having verbal responses different from those usually reinforced by the parent and, also, durable verbal behavior that will continue to be emitted under the intermittent reinforcement that is likely to occur. If the child's repertoire is durable and extensive enough so that the verbal response may be repeated several times and supplemented by auxiliary behavior, the child has a greater chance of affecting the new person or of being shaped by him. Similarly with other kinds of social behavior. The wider the range of behavior and the greater the disposition to emit it the more likely that the child's performance will be within the range of responses potentially reinforcible by the new environment.

For a stimulus to acquire control over behavior, the child must first emit behavior in the presence of the stimulus. Consider, for example, the performance of a child at a children's party at which there are lots of toys and games, such as bicycles, swings, and so forth. The likelihood of the child's behavior coming under the control of any of the other children as reinforcers is minimal if the new environment suppresses or makes the child's entire repertoire unavailable because it is a novel stimulus and is an occasion on which the child's behavior has never been reinforced. If the behavior of playing with a swing or riding a tricycle is sufficiently strong that it may be emitted even under the adverse conditions of the very strange party environment, then the simple emission of the previously developed behavior provides a situation under which other children at the party may potentially reinforce or otherwise affect the child's repertoire. Simply the acts of eating cake, candy, or ice cream, or picking up a toy put some of the child's behavior under the control of the new environment. Each new performance which can potentially occur at the party provides a basis for the child's reinforcing some behavior of other children at the party or of his coming under the control of the other children's reinforcers. On the other hand, a sudden exposure to a new environment with a weak and narrow repertoire may produce a severe behavioral deficit. In any case, the child will be much less disposed to go to the party if he had behaved unsuccessfully in the new environment. This lower disposition to attend and engage in the party would in turn make it less likely that the child will emit behavior that would be reinforced in the party environment.

Adaptation

The emotional and elicited autonomic effects of novel environments may also interfere with a child's performances. Adaptation to new environments occurs with gradual exposure. A sudden exposure to a new environment will produce gross emotional and autonomic responses which will in turn interfere with, or even completely suppress, the emission of possible operant behavior potentially reinforcible by the new environment. The rate at which the child is exposed to the new environments will determine the magnitude of disturbance. Exposure to a new environment and adaptation of the emotional responses do not necessarily create the potential basis for responding, however. A repertoire that will make contact with the new environment is also necessary.

Amount of Prior Nonreinforcement

The more closely controlled the child's performances are by specific stimuli, the more likely a sudden shift in the environment will produce a cessation of responding. For example, the child receiving minimal care from a parent probably will be less affected by a sudden shift in environment than a child closely affected and controlled by parental response. It is paradoxical that the parent who responds sensitively to the child's performance may be potentially weakening it more than the parent who exerts little control over the child. It is the alternate reinforcement and nonreinforcement that place the child's behavior narrowly under the control of very specific stimuli so that it is much more vulnerable to sudden changes. The range of stimuli in whose presence the child's behavior goes unreinforced will determine the narrowness of the stimulus control.

Cumulative Effects of a Behavioral Deficit

The continuous development of more and more complex forms of a child's behavior is normally achieved because the parents and community approximate the required performances. At each stage of the child's development, the community reinforces the child's current repertoire even though it is more disposed to react to small increments in the child's performance in the direction of the required complex performances. Should any of the above processes produce a deficit in performance or an arrest in the development of the child's performance, further development of a repertoire would depend upon the community's relaxing its requirements and reinforcing performances in an older child that it normally accepts only from a younger one. Ordinarily, the reinforcing practices of the community are based on the chronological age and physical development of the child.

Only between the ages of 1½ to 4 years does the parent have sufficient control of the child to weaken his performance to the degree seen in infantile autism. This is a critical period in the child's development during which his behavior is especially susceptible to extinction, because the traditional social pattern in the usual family restricts the child's experience to one or two parents. Before the age of 1½, the child has few performances with which the parent will interfere or that have important effects on the parent. Much of the infant's behavior is maintained by simple and direct effects on its environment. As the child approaches 2 years, the rapid development of a behavioral repertoire, particularly social and verbal behavior, makes possible extinction and other forms of weakening. The effectiveness of the parental environment in weakening the child's repertoire depends upon the availability of concurrent audiences for the child's behavior. In general, the 2-year-old child is limited to the home and comes into increasing contact with other environments as he grows older, perhaps reaching a maximum at school age. The presence of an older sibling might appear to preempt the possibility of a sufficient degree of isolation to account for an aversive behavioral deficit. A sibling could provide an alternative to the parent as a reinforcing environment. The behavioral or functional influence of a sibling would depend on the amount and nature of interaction between the children. For example, an older child might possibly completely avoid the younger one or tend to have the same patterns of reaction as the parent. In many cases, the older sibling has playmates outside the home to the complete exclusion of the younger child. The older sibling, in many circumstances, punishes as well as extinguishes the younger child for any attempted participation in his play. There are very few facts as to the exact nature of the interactions in most cases.

The parent as the sole maintainer of the child's behavior is perhaps even more likely when the child is raised in a rural or isolated community, and perhaps with one of the parents largely absent. The above analysis suggests that a survey of severely autistic children would, in general, show them to be first-born children; or, if other siblings were available, they would have provided little interaction with the child. It also suggests that the child would be raised in a house physically or socially isolated from other families or children such that there were no alternative social environments that could provide reinforcement for the child's behavior. When the child was exposed to both parents, it would be expected that both parents were consistent in their nonreinforcement of the child's performances.

AVERSIVE CONTROL AND PUNISHMENT

It has been possible to describe conditions which might produce major behavioral deficits without dealing with punishment or aversive control. A similar account might present a functional analysis of how performance deficits might occur as a result of aversive control. Many writers have already described some of these factors by extending general principles of aversive control to human behavior (7, 8, 11). For the purposes of the analysis presented in this paper, I should like to restrict the discussion of aversive control to its relation to positive reinforcement. Much of human aversive control is carried out by discontinuing or withdrawing reinforcement (3, 10). For example, a frown or criticism may function as an aversive stimulus because these are occasions on which reinforcements are less likely to occur. Even when corporal punishment is given, it is not clear as to whether the resulting effect on the child's behavior is due to a slap or to the lower inclination of a punishing parent to reinforce. Most parents who spank a child will be indisposed to act favorably toward the child for some period of time subsequently. As a result, one major by-product of frequent punishment may be a larger order of interference with the child's normal repertoire along the lines of the positive reinforcement deficits described above.

The obvious effectiveness of punishment in some kinds of human control appears to contradict experimental findings with animals which show punishment to have only a temporary effect on behavior (1, 2, 9). The role of positive reinforcement factors helps resolve the dilemma. The effectiveness of punishment depends on how strongly the punished behavior is maintained by positive reinforcement. The apparent effectiveness of punishment in the control of children may occur when weak repertoires are punished or when the punishment indirectly produces extinction. Most animal experiments using electric shock as an aversive stimulus have used strongly maintained positively reinforced operant behavior as the base-line performance to be punished. The aversive control might be more effective when the performances to be punished are less strongly maintained.

CONCLUSION

As might be expected from the relatively low frequency of infantile autism, the combination of circumstances hypothesized above would occur only rarely. The above hypothesis provides a framework for investigating the circumstances surrounding the development of the autistic child. All of the variables that might weaken the behavior of a child are directly or potentially observable. The data required are the actual parental and child performances and their specific effects on each other, rather than global statements such as dependency, hostility, or socialization. Not all of the factors responsible for a child's performance may be present currently. Using retrospective accounts, however, makes it difficult to determine the actual correspondence between the verbal statements of the parent and their actual practices in raising the child. The alternatives are, first, an objective assessment of the child's repertoire in a wide enough range of environments so as to allow an assessment of the nature of the environmental control of the child's current behavior; and, second, actual home observations of the specific social consequences of the child's performances early in the development of the disease.

The same kind of functional analysis can be made for the performance of the adult psychotic although the specific deficits observed in autistic children and their manner of occurrence may not be relevant. In particular, the analysis of the adult's behavior would be more concerned with the factors which weaken behavior already in the repertoire rather than the development of new repertoires as with the analysis of the autistic child's behavior. Maintaining already-established behavior is more at issue in the adult than the initial development of a performance as in the case of the child (3).

REFERENCES

1. AZRIN, N. H. Punishment and recovery during fixed-ratio performance. *J. exp. anal. Behav.*, 1959, 2, 301-305.
2. ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, 57, No. 3 (Whole No. 623).
3. FERSTER, C. B. Reinforcement and punishment in the control of human behavior by social agencies. *Psychiat. Res. Repts. 10,* 1958, 12, 101-118.
4. FERSTER, C. B., & DEMYER, M. K. A method for the experimental analysis of the behavior of autistic children. *Amer. J. Orthopsychiat.*, in press.
5. FERSTER, C. B., & DEMYER, M. K. The development of performances in autistic children in an automatically controlled environment. *J. chronic Dis.*, 1961, 13, 312-345.
6. KANNER, L. Early infantile autism. *J. Pediat.*, 1944, 25, 211-217.
7. MILLER, N. E., & DOLLARD, S. *Personality and psychotherapy.* McGraw-Hill, 1950.
8. ROTTER, JULIAN. *Social learning and clinical psychology.* Prentice-Hall, 1954.

9. Skinner, B. F. *The behavior of organisms.* Appleton-Century-Crofts, 1938.

10. Skinner, B. F. *Science and human behavior.* Macmillan, 1953.

11. Skinner, B. F. Some contributions of an experimental analysis of behavior to psychology as a whole. *Amer. Psychologist,* 1953, 8, 69-78.

12. Skinner, B. F. *Verbal behavior.* Appleton-Century-Crofts, 1957.

COMPARISON OF DATA OBTAINED FROM MOTHERS AND FATHERS ON CHILDREARING PRACTICES AND THEIR RELATION TO CHILD AGGRESSION[1]

LEONARD D. ERON,* THOMAS J BANTA,[2]
LEOPOLD O. WALDER, *and* JEROME H. LAULICHT[3]

Rip Van Winkle Foundation

In reviewing the literature having to do with behavior of children as it is related to behavior and attitudes of parents, one is struck by two methodological biases which do not permit a clear-cut evaluation of the substantive findings of a majority of these studies. The first has to do with the disregard of the father both as an important socializing influence and as an informant. A search of the literature between 1929 and 1956 revealed 160 publications dealing with mother-child relationships but only 11 with father-child relationships (10). Since that time there have been a few more studies which have considered the importance of father-centered variables, e.g., Bronfenbrenner (2), Miller and Swanson (9), Sears, Maccoby, and Levin (11), Sears, Pintler, and Sears (12), but information about fathers in these studies usually has been obtained second-hand, either from the mother or from the child himself. This introduces the second methodological bias which is the contamination that is likely when assessments of both parental and child behavior (presumed antecedent and consequent) are made by the same individual. A study in identification by Levin and Sears (8), as well as the earlier Sears' studies (12, 13), are not subject to this error, although they are to the first. When both the predictor and criterion are obtained from the same subject, then response set cannot be eliminated from the possible interpretations of the results. In addition to contributing important information which has usually been lacking in childrearing studies, the use of both mother and father as informants could serve as one check on this source of error. The optimal procedure, of course, is to have completely

---

* Mental Health Research Center, Rip Van Winkle Foundation, 886 Columbia Street, Hudson, New York.

[1] This is a revised version of a paper delivered at the Eastern Psychological Association meetings, New York City, April 16, 1960. The research has been supported by grant M-1726 from the United States Public Health Service. Thanks for cooperation and support are due to W. Richards Bonneau, Principal, Ockawamick Central School; Robert Cullen, Principal, Claverack Union Free School; George Pike, Elementary Supervisor, Hudson Public Schools; Jack Roosa, Principal, Greenport School; and the following third-grade teachers: Mrs. Kathelyn Caswell, Mrs. Grace Clapp, Mrs. Rose Healy, Mrs. Catherine McNamara, Mrs. Patricia McNamara, and Mrs. Winifred Rockefeller. We should also like to thank all the children and their parents, who of course must remain anonymous, for their excellent cooperation.

[2] Now at University of Wisconsin.

[3] Now at Berkshire Industrial Farm, Canaan, New York.

Child Develpm., 1961, 32, 457-472.

independent observations of the child and individual measures obtained from both parents. A few recent studies have approximated this goal. Davidson *et al.* (5) administered anxiety questionnaires to a group of children; they then interviewed the mothers *vis a vis,* but merely left a check list of personality characteristics of the child to be completed by the father. Both parents were interviewed by Peterson *et al.* (10), but they did not have independent observations of the child's behavior; although they partially overcame this type of contamination by comparing the attitudes of one parent with judgments of the child's behaviors based on reports of the other parent. Despite these limitations on clear-cut interpretation, both of these studies showed differential results when predicting to child behavior from either mother or father interviews. In the study here reported it has been possible to interview fathers and mothers separately and to obtain independent measures of the children from other informants. It is with the relations among items from these three sources of information that this paper is concerned.

SUBJECTS AND PROCEDURES

A group of 158 children, the total population of six third-grade classes from divergent socioeconomic areas in a semirural county, served as a pool of subjects. Sixty children were drawn from this group on the basis of their scores on an aggression sociometric (described below) administered in the classroom. Thus, the subjects included the 10 most aggressive boys in this sample, the 10 least aggressive boys, 10 boys in the middle range, and three similar groups of girls. Both parents (where available) of each of these children were contacted for personal, individual interviews. The families of only three of these preselected children refused to participate, and they were replaced by subjects with very comparable scores. It was thus possible to interview all of the mothers and 50 of the 58 available fathers. The interviews were conducted by six experienced interviewers[4] either in the subject's home or the Foundation offices. At the time of the interview, the interviewers had no knowledge of the aggression classification of the subject.

MEASURES

Aggression Index

This is a Guess-Who technique in which every child in a class rates everybody else on a selected series of 22 specific aggression and four aggression anxiety items. These 26 items were what remained after a process of "sifting and winnowing" from an original pool of 1000 short behavioral descriptions which had originally been collected from the literature and by interviewing experts:

---

[4] Thanks are due to Dr. Marjorie Collins and Mrs. Madeline Eron, psychologists, and Miss Mary Lawrence, psychiatric social worker, who assisted three of the authors (LE, JL, and LW) in obtaining these interviews.

THE AGGRESSION INDEX

(The labels in parentheses refer to type of aggression.)

Cluster 1

1. Who is a pest? (Indirect)
2. Who does not obey the teacher? (Unclassified)
3. Who takes the teacher's things without permission? (Acquisitive)
4. Who is always getting into trouble? (Unclassified)
5. Who tattles to the teacher? (Indirect)
6. Who is rude to the teacher? (Verbal)

Cluster 2

7. Who starts a fight over nothing? (Physical)
8. Who says mean things? (Verbal)
9. Who makes it hard for children to get things done? (Indirect)
10. Who pushes or shoves children? (Physical)
11. Who does things that bother others? (Indirect)

Cluster 3

12. Who forgets to return borrowed things? (Acquisitive)
13. Who often says "Give me that"? (Acquisitive)
14. Who makes marks on the desk? (Physical, against property)
15. Who takes other children's things without asking? (Acquisitive)

Individual Aggression Items

16. Who will always fight back if someone else hits them first? (Physical)
17. Who gives dirty looks or sticks out their tongue at other children? (Verbal)
18. Who complains to the teacher when she tells them what to do? (Verbal)
19. Who grabs things from other children? (Acquisitive)
20. Who uses bad words when another child bothers them? (Verbal)
21. Who gets very, very mad at times? (Unclassified)
22. Who makes up stories and lies to get other children into trouble? (Indirect)

Anxiety about Aggression Items

23. Who is always polite? (Unclassified)
24. Who will never fight even when picked on? (Physical)
25. Who will never argue even when they are right? (Verbal)
26. Who says, "Excuse me," even when they have not done anything bad? (Unclassified)

Rejection by Peers

27. Who are the children that you wish were not in your class at all?

Popularity with Peers

28. Who are the children that you would like to have for your best friends?
29. Who do you know best of all?
30. Who would you like to sit next to in class?
31. Who would you like to play with?

They represented the final distillation of a long series of preliminary studies including armchair judgments by experts as to type of aggression involved, pilot studies with groups of 8-year-olds, and an extended tryout of two alternate forms on 974 third graders, with subsequent cluster and item analyses. The purpose was to find items on which the children, as judges, could agree with each other as to who of their classmates did what. Also included in this sociometric were one rejection and four popularity items. The procedure involved giving each child a booklet containing one page for each question. The names of all children in the class were listed on each page, and subjects were instructed to cross out the names of those children who acted in the way described by the particular question for that page which was read aloud to them by the examiner. The score for each child was based on the number of judges choosing him as someone who fitted a particular behavioral description. Thus, if 10 members of a class crossed out Johnny Jones' name as someone who said mean things, Johnny's raw score was 10 for that item. The raw scores were converted into percentages to correct for differences in class size.

The reliability of this instrument, based on interjudge agreement, had been shown to be excellent, with minimum shrinkage on cross-validation. A cluster analysis isolated three sets of items, one concerned with aggression against teacher, one with aggression against peer, and one having to do with a particular type of aggression which was called acquisitive (*see* Table 1). Also included in this measure which was derived from the two alternate forms were a number of individual items, each with good reliability (at least .70). These had been retained in order to tap different types of aggression, different objects, and different provocation levels.

TABLE 1

CLUSTER RELIABILITIES

	Cluster	*Median Reliability**	*Cross Validation*
Teacher as object	1	.92	.86
Peer as object	2	.87	.77
Acquisitive	3	.85	.80

* The median reliability is based on 12 different groups (six classes, each divided by sex into two groups). Typically, the mean is higher than the median.

There was no relation between score on the aggression measure and amount of interpersonal activity the child engages in as rated by the teacher (r ranged from $+.22$ to $-.38$ in six classes with median r of .00) nor with the position of a child's name on the list (r ranged from $+.19$ to $-.15$ with median r of .00). As for the distribution of the aggression scores among the subjects, it was found that very few children were highly aggressive; however, the vast majority were moderately aggressive and girls had lower scores than boys. The more popular children were less aggressive, but even the most popular children were rated as aggressive in some ways. On the cluster having to do with aggression towards the teacher, identically shaped distributions of scores (J curves) were obtained for boys and girls. This type of distribution would be expected in most kinds of nonconforming behavior, of which aggression towards teachers is certainly an example.

In addition, Cureton's r_{rb} (4) of .84, uncorrected for attenuation, was obtained between teacher's and peers' rating on this cluster. It is reasonable to assume that the high reliability coefficients themselves indicated validity because the raters were probably basing their judgments on observations of the same specific behaviors (3, 6). The shape of the distribution curves and its consistency from class to class and between boys and girls were further evidence of validity, as were the predicted relations into which the aggression and aggression anxiety items entered with the other classroom measures: popularity-rejection, IQ, and teacher's ratings of aggression (*see* Table 2). There was also some clinical evidence of validity since a few of the subjects had been referred to the Clinic Guidance Service prior or subsequent to the administration, and in each case, the children who were shown to be overly aggressive scored high on the sociometric.

TABLE 2

INTERCORRELATIONS AMONG CLASSROOM MEASURES*

Other Measure	*Aggression Measure*	*Aggression Anxiety*
Rejection	.73	—.46
Popularity	—.38	.48
Teacher rating on all items	.63	—.34
IQ	—.17	.20

* A correlation of .23 is significant at the .01 level of confidence.

Parent Interview

This is an objective, almost entirely precoded interview which takes about 1½ hours to administer. Its purpose is to get at the sociocultural and psychological antecedents of aggression as they are mediated by child-parent interaction. Included in this interview, which was pretested on 200 mothers and fathers, are a number of variables suggested as important by both general behavior theory and clinical hunches. It was felt that a profitable way to look at the differential distribution of behaviors which exist in varying degree in different segments of society was through the learning environments provided by these subcultures. Thus, the pertinent variables had to do with rewards and punishments for aggressive behavior, standards of aggressive behavior, role models, identification, and presumed frustrating antecedents to such behavior. These variables are listed below, with definitions and examples of items to tap them:

PARENT INTERVIEW VARIABLES

1. *Approval of Aggression*—Evaluative standards for aggressive behavior stated in terms of approval or disapproval of specific items of aggressive behavior which appear in the school aggression measure. Example: Suppose

NAME said mean things to another child. Would you: 0. strongly disapprove, 1. mildly disapprove, 2. not care (and don't know), 3. mildly approve, 4. strongly approve.

2. *Confessing by Child*—Extent to which a child behaves as if he were monitoring his own behavior in a way he thinks a socializing agent would. These items are closed-end versions of two questions by Sears, Maccoby, and Levin (11). Example: When NAME has done something naughty and you haven't seen him do it, does he come and tell you about it without your having to ask him? 4. all the time, 3. most of the time, 2. some of the time, 1. almost never, 0. never.

3. *Dependence Avoidance of Child*—Inability or unwillingness of the child to accept help or rely on others. Example: Does NAME seem embarrassed when you take his part? 0. no, 1. sometimes (and don't know), 2. yes.

4. *Ethnicity* (generational level)—Number of generations in which parents' forebears have lived in the United States.

5, 6. *Father's and Mother's Aggression*—Tendency of the parent to display aggressive behavior in situations which often elicit aggressive responses. Each parent rates himself and is rated by the other parent on each of these situations. Example: Suppose you are driving a new car and get into an accident which is clearly the other driver's fault. Would you show your anger if: 1. he says, "What's the matter, can't you drive?" 2. he apologizes? 3. he says, "I'm not going to say anything until I see a lawyer"? 4. he laughs it off?

7. *Home Aggression of Child*—Frequency of acts whose goal response is injury to another object. Example: How often does NAME say mean things to another child? 0. never, 1. rarely, 2. occasionally, 3. pretty often, 4. frequently, 5. daily. These alternatives were specifically defined in terms of frequency in time and typed on cards which were handed to the respondent.

8. *Lack of Social Participation*—Degree of participation of respondent in formal and informal social relationships. Example: About how many times in the past year have you attended meetings or affairs of any *local* organizations, societies or clubs? 0. 13+, 1. 7-12, 2. 4-6, 3. 1-3, 4. none.

9. *Nonrecognition of Child's Needs*—An aspect of nonnurturant behavior. (For other aspect of nonnurturant behavior see *Punishment for Nurturance Signals* below.) Example: Do you usually have time so that NAME can talk to you about things that interest him? 0. yes, 1. no (and don't know).

10. *Parental Aspirations for Child*—Level of education parent hopes child will attain. Example: How much education do you expect NAME to get? 1. high school + specialized training or college, 2. high school graduate or less.

11. *Parental Aspirations for Self*—Example: When you left school what particular kind of occupation or life work was it your ambition to reach some day? Aspirations for: 1. professional status, 2. minor profession, small business or farm owner, 3. skilled worker trades, 4. semiskilled or unskilled occupations.

12. *Parental Disharmony*—The extent of disharmony in the home as measured by disagreement about various specific matters of importance in a family; items dealing with arguments between husband and wife, and presumptive evidence such as separation, divorce, amount of time spent together, etc. Example: Are you satisfied with how your SPOUSE handles money? 0. yes, 1. sometimes (and don't know), 2. no.

13. *Parental Rejection*—The number of changes in the child's behavior (aggression excluded) and characteristics desired by the socializing agent. The parent is considered to be accepting when he indicates that his needs are

satisfied by the child: "I like you the way you are." Example: Do you think NAME wastes too much time? 2. yes, 1. sometimes (and don't know), 0. no.

14. *Parental Restrictiveness*—Extent to which the child defines behaviors which are proper for him to perform rather than the agent defining proper behaviors for the child. Restrictiveness refers to the amount of control exercised by the agent over the child. Example: Do you make NAME finish up everything he is served as mealtime? 2. yes, 1. sometimes (and don't know), 0. no.

15. *Punishment for Aggression*—Rewards and punishments of various *intensities* administered by socializing agents contingent upon the child's aggressive behavior. Example: What do you usually do when NAME is rude to you? (Verbatim response and probes recorded, subsequently rated by three judges on a scale from 1 to 7.) 1. rewarding aggression, 2. don't do anything, 3-7. mild to severe punishment for aggression.

16. *Punishment for Dependency*—Rewards and punishments of various intensities administered by socializing agent when child asks for help. Example: What do you usually do when NAME asks for help? (Each response was rated by three judges on a scale from 1 to 4): 1. giving help, to 4. punishing the child.

17. *Punishment for Nurturance Signals*—Rewards and punishment of various intensities administered to child by socializing agent in situations which might tend to lead to nurturant behavior on part of agent. Example: What do you usually do when NAME is afraid? (Each response was rated by three judges on a scale from 1 to 4): 1. giving nurturance, to 4. punishing the child.

18. *Residential Mobility*—A measure of the number of times the child changed residence and thereby had to change schools and/or find new friends. Example: How often has moving meant that NAME had to find new friends? 0. none, 1. once, 2. 2-3 times, 3. 4+.

19. *Rural Background*—Population size of geographic area in which respondent was born and grew up. Example: Where did the family live when you were born—on a farm, or in a village, town, small city, medium-sized city, or big city? 1. big city (500,000+), 2. medium-sized city (100,000-500,000), 3. small city (10,000-100,000), 4. town (1,000-10,000), 5. village (under 1,000), 6. farm.

20, 21. *Shame, at home and out of home*—Tendency to punish in public assessed by items involving different kinds of punishments and different publics. Example: Suppose NAME was naughty and you felt he deserved a scolding. Would you do it when: (1) your SPOUSE and other children could hear it? (2) one of NAME's friends could hear? (3) one of your close friends or relatives could hear? (4) a neighbor or acquaintance could hear? (5) you were in public and someone else might hear? 0. no. 1. sometimes (and don't know), 2. yes. The first item constitutes a separate measure, punishing the child in front of his family which defines home shame.

22. *Social Isolation of Child*—Frequency and type of contacts with peers outside of school. Example: About how many children of NAME's age live in the neighborhood? Would you say about one or two, three to five, or more than five? 3. none, 2. 1-2, 1. 3-5, 0. 5+.

No definition was accepted for these variables until items could be written which were judged by a number of experts to fit the definition. Emphasis was placed on the necessity for distinctive definitions and measures of each concept in order that hypotheses about how they relate to overt aggression in school could be tested both singly and in combination. The median number of items for each variable was 10, with a range from 1 to 20.

Before the relations of mother and father interview variables to each other and to the independently obtained criterion measure were tested, the homogeneity of the variables was increased by eliminating all items which correlated less than .30 with the total scale score of which it was a part. These item-total correlations were done separately for mothers and fathers since it was felt that there would be some questions more appropriately asked of fathers and others of mothers. For 14 of the 22 variables under consideration, overlap of questions between mother and father was complete. Of the remaining eight, subsequent analyses on six of them showed there was little difference in results whether or not items completely overlapped (1). It turns out that it makes little difference in this report whether scores are calculated on the basis of the refined scales with incomplete overlap between mother and father questions or on scales with identical questions for both groups.

Results and Discussion

Once homogeneous sets of items had been derived, a number of correlational analyses were done. First, scores on each variable from mother's interview were correlated with scores on similar variables in the father's interview. As is shown in Table 3, of 22 correlations, only 10 were significantly better than zero ($p < .05$). It is obvious that mothers' and fathers' responses cannot be substituted for one another, at least in this childrearing questionnaire.

An examination of Table 3 reveals that the higher correlations generally are for those variables which have to do with the respondent himself, his

TABLE 3

CORRELATIONS BETWEEN MOTHERS AND FATHERS ON INTERVIEW VARIABLES*

Variable	*r*	*Variable*	*r*
Residential Mobility	.91	Confessing by Child	.25
Parental Rejection	.64	Ethnicity (Generational Level)	.21
Lack of Social Participation	.63	Nonrecognition of Child's Needs	.20
Parental Aspirations for Child	.52	Shame Out of Home	.16
Parental Disharmony	.46	Mother's Aggression	.15
Social Isolation of Child	.42	Punishment for Nurturance Signals	.13
Punishment for Aggression	.42	Approval of Home Aggression	.09
Parental Restrictiveness	.38	Parental Aspirations for Self	.07
Father's Aggression	.35	Shame at Home	—.06
Rural Background	.32	Child's Dependence Avoidance	—.05
Home Aggression of Child	.25	Punishment for Dependency	—.04

* $r_{.05} = .28$.

own attitudes and feelings; the lower correlations, for the most part, are for those variables having to do with the child's behavior or the parent's interaction with the child. This reveals something about marital choice perhaps (likes attract), but it also shows that mother and father cannot agree in reporting their perceptions of their own child's behavior. However, on the one really objective item that both parents rated, residential mobility, the correlation was .91. Thus, it is not that the parents cannot render reliable information, but perhaps mother and father each observe and react to children differently, and the observations and reactions of each must be taken into account to get a complete picture of parental socialization influences on child behavior.

The same conclusion can be drawn from the next analysis which was the intercorrelation of each interview variable with the other, as well as with the independently obtained classroom measures, done separately for mothers and fathers. For any one variable, all correlations significant at or beyond the .05 level for either mother or father are reported. These appear in Table 4. Although the correlation between mothers' and fathers' scores on any one variable may be better than zero, their scores on that particular variable do not relate to those on other variables in the same way. Thus, for example, the correlation between mother's and father's rating of rejection is .64; however, while mother's rejection score relates significantly to all the classroom measures, father's rejection relates to only one of them. On the other hand, father's rejection relates to a number of other variables in his interview while mother's rejection relates to none of these.

The opposite is true for ethnicity or generational level, a classification of parents on how long ago their ancestors came to this country (on which

TABLE 4

SOME CORRELATIONS AMONG SCALES

Variable	*Mother* $N=59$ $r_{.05}=.26$		*Father* $N=50$ $r_{.05}=.28$
A. *Parental Rejection* (r mother-father = .64)			
School Aggression	.40		.31
Aggression Anxiety	—.37		—.17
Popularity with Peers	—.29		—.12
Peer Rejection	.35		.20
Home Aggression of Child	.11	*	.49
Confessing by Child	—.07	**	—.58
Parental Disharmony	.10	*	.42
Parent's Aggression	—.03	*	.32
Rural Background	—.12		—.37
Punishment for Dependency	.07		.36
Punishment for Aggression	.20		.44

(*continued on next page*)

TABLE 4 (*continued*)

SOME CORRELATIONS AMONG SCALES

Variable	Mother $N=59$ $r_{.05}=.26$		Father $N=50$ $r_{.05}=.28$
B. *Ethnicity* (generational level)			
(r mother-father = .21)			
School Aggression	.00		.26
Popularity with Peers	.08	*	—.30
Peer Rejection	—.03	**	.41
Home Aggression of Child	.11	**	—.29
Lack of Social Participation	.06		.28
C. *Parent's Aggression*			
(r for mother's aggression = .15; r for father's aggression = .35)			
Home Aggression of Child	—.05	**	.36
Parental Rejection	—.03	*	.32
Spouse's Aggression	.24		.29
Punishment for Dependency	—.10	**	.33
Nonrecognition of Child's Needs	—.37	*	—.02
D. *Punishment for Aggression*			
(r mother-father = .42)			
School Aggression	.03		.31
Home Aggression of Child	.47		.46
Confessing by Child	—.17		—.35
Parental Disharmony	.28		.18
Parental Rejection	.20		.44
Parental Restrictiveness	.42		.28
Approval of Home Aggression	.00		—.28
Rural Background	—.26		—.16
Shame at Home	.16		.32
Child's Dependence Avoidance	.00	*	.38
Nonrecognition of Child's Needs	—.35		—.14
E. *Confessing by Child*			
(r mother-father = .25)			
School Aggression	—.21		—.31
Peer Rejection	—.10		—.29
Home Aggression of Child	—.28		—.07
Residential Mobility	.04	*	—.32
Parental Rejection	—.07	**	—.58
Approval of Home Aggression	—.32		—.14
Rural Background	.12		.28
Lack of Social Participation	.04	*	—.34
Punishment for Dependency	.10	*	—.28
Punishment for Aggression	—.17		—.35
Punishment for Nurturance	—.04		—.28

* Difference between r for mother and father significant at .10 level.

** Difference between r for mother and father significant at .05 level.

there is no significant relation between mother and father). Here it is the father's score that is discriminating. Whereas mother's ethnicity was not significantly related to any other variable, father's ethnicity was significantly related to rejection by peers and home aggression of child, among other things. The relation between father's ethnicity and child's behavior is especially significant when fathers and sons are paired. For example, father's ethnicity related $+.53$ ($p < .001$) to son's school aggression and $-.53$ to his home aggression. One possible interpretation would be that ethnic fathers do not tolerate aggression in the home, and thus the tendency on the part of the son to aggress is displaced to the school context. Another possibility is that ethnic fathers may provide a more aggressive role model by their punitive behavior at home. At any rate, it would seem that father's ethnicity has a more important influence on the child's behavior and the reaction of others to him than does mother's ethnicity, at least for boys (correlation between father's ethnicity and daughter's aggression fails to reach significance at the .05 level; indeed, the only significant correlations for ethnicity of girls' fathers are with social participation [$-.40$] and sanctions for nurturance [$-.42$]).

The likelihood of contaminated data which is present when both predictor and criterion measures are obtained from the same informant, as is often done in childrearing studies, is seen in the relation of the parent's self-rating of aggression to other variables. This measure is not related to school aggression or any other independently obtained measures but is related, at least for fathers, to ratings of the child's home aggression, as well as spouse's aggression and rejection. The relation of parent's aggression to rejection is especially marked for father-daughter pairs (.52). Furthermore, those fathers who score high on rejection of their daughters also tend to rate their wives as aggressive (.49). Thus, those respondents who are willing to admit that they themselves are aggressive will also tend to say that wife and daughter are aggressive and that daughter does not measure up to their standards. Since father's self-rating of aggression related only to his own rating of child's aggression and not at all to the independently obtained rating, response set is certainly a compelling consideration.

Again, although both mother's and father's punishments for aggressive behavior are related to the frequency of aggression at home, in each case both these sets of items are reported by the same informant, and response set must also be considered as a reason for this relation. Only father's punishment is significantly related to the independently obtained measure of school aggression. This relation holds especially for father-son pairs ($r = .46$). Thus, it seems that what the father does in response to the child's home aggression is more likely to influence aggressive behavior of the son than what the mother does. The same results obtained when each of the 13 items of aggressive behavior in school which comprised the aggression score from the parent's interview were considered separately. It was still the more punitive father who had the more aggressive child in school. The results for mother were in the same direction, but none of the correlations between punishment by the mother for a specific behavior and the appearance of that behavior in school was significant. The importance of the father as a role model for aggressive behavior is here strongly suggested. Levin and Sears (8) have stressed the importance of the father as a model. In a study of fantasy aggression they found that boys (attending nursery school) who were punished by their fathers and who were also identified with their fathers tended to be more aggressive in the doll play situation than those not so identified or punished. Their measure of identification was a rating of superego formation based on four open-ended questions having to do with confessing and denying by the child of proscribed acts he was known to have committed. These questions were asked of mothers only.

In the present study two of these four questions were adapted to a closed-end version, thus forming the scale of confessing. They were asked of both mothers and fathers with differing results. It was found that indeed, confessing behavior of the child as rated by the father was related to the independently obtained measure of school aggression, but in the opposite direction from the Levin and Sears results and with these results holding for both boys and girls (—.39 for girls; —.37 for boys). However, the relation between confessing and school aggression was not significant either for mother-son or mother-daughter pairs. The same holds true for the relation between confessing and punishment for aggression. There is a significant relation for fathers but not for mothers, especially for father-son pairs (—.53). The important feature in the aggression-confessing-punishment relation may indeed be the punitiveness which would lend support to the role model hypothesis. At any rate, these results are similar to those of Levin and Sears in stressing the importance of the father, especially for boys. The difference in direction of the results is likely a function of the difference in the nature of the criterion, in one case real-life aggression as rated by the child's peers, and in the second case, fantasy aggression, as observed in doll play by the experimenters. In the latter case, the child, who is rated by the experimenter, is actually playing the role of the "daddy" doll; in the former he is rated by his peers as he, the child himself, behaves in the natural setting. Also accounting for the variation may be the difference in age level, preschool vs. third grade, as well as the different operations used in measuring the variable, confessing.[5]

Discussion thus far has dealt with correlations between measures obtained from the same informant. When cross-over correlations are considered, e.g., the correlation of each mother score with all other father scores, there is considerable shrinkage in size of correlation. For example, the correlation between father's rating of rejection and father's rating of frequency of aggression is .46, but the correlation between father's rating of rejection and mother's rating of frequency of aggression is only .15.[6] A separate analysis was also done in this manner for each of the 13 individual items of aggressive behavior which go to make up the composite child's home aggression score. A significant relation between mother's punishment and mother's rating of child's home aggression was found on five items, between mother's punishment and father's rating of child's home aggression on only two. In the same way, father's rating of punishment correlates with father's rating of child's home aggression significantly on four of the 13 items and with mother's rating of child's home aggression on none of them. This consistent drop, when going from correlations between predictor and criterion obtained from the same individual to correlations between predictor and criterion each independently obtained, reinforces suspicions of contamination in those studies which use only the mother as informant about both childrearing practices and child behavior.

---

[5] Levin (7) has stated that the use of superego development as an index to identification, "assumption of adult role models by the child," may be valid for young children but that by the age of 8 years role modeling and superego functioning quite likely have differentiated into separate behavior patterns.

[6] For an extended discussion of the results of this cross-over analysis, see paper by Banta *et al.* (1).

In general, it would seem that mothers and fathers observe, sanction, and report their children's behavior from different vantage points. Which one should we accept, and which ignore? Traditionally, childrearing studies have utilized information obtained only from the mother. Our results show that indeed mothers may be more discriminating observers in some areas; however, in others, fathers give us more consistent results. As noted above, mother's rejection related to all the classroom measures and none of the interview measures, while the reverse was true for fathers. Father's punishment for aggression and father's ethnicity, on the other hand, related to school aggression, while neither mother's punishment nor ethnicity did. Father's ethnicity was also related to rejection by peers as well as to a number of variables within the interview, while mother's ethnicity was related to no other variable. Moreover, these relations were more marked for father-son than for father-daughter.

It seemed obvious that both fathers' and mothers' reports had to be taken into account. Therefore, an attempt was made to predict from combined mother and father scores on various interview variables to the classroom aggression scores, and indeed it was found that variables which did not relate when predictions were made from the scores of only one parent did show a relation to the criterion when scores of both parents were considered. On each variable the mothers and fathers were split into high and low groups at the median score for their respective samples. The median classroom aggression scores for the four cells containing subjects classified according to whether mother and father were each or both high or low on the specific variable appear in Table 5.

This combination of mothers' and fathers' scores clarifies the relation between father's ethnicity and child's school aggression. If it is only when there is a high ethnic father and low ethnic mother that the aggression score is very much elevated. When both mother and father are ethnics or both are nonethnics, the aggression score is near the median for the entire group. On the other hand, when the mother is high ethnic and the father is low ethnic, the median school aggression score is noticeably low. This suggests that there may indeed be a conflict of cultural values behind the obtained

Table 5

MEDIAN SCHOOL AGGRESSION SCORES OF SUBJECTS BY COMBINED PARENT POSITION ABOVE AND BELOW MEDIAN ON SOME INTERVIEW VARIABLES

	Hi Father		Lo Father	
Variable	*Hi Mother*	*Lo Mother*	*Hi Mother*	*Lo Mother*
Ethnicity	15.5 (14)	69 (8)	6 (9)	16 (19)
Confessing by Child	5 (15)	22.5 (8)	16 (13)	25.5 (14)
Parental Disharmony	15.5 (16)	14 (8)	45 (8)	8.5 (18)
Parental Rejection	36 (17)	3.5 (8)	13 (9)	4.5 (16)
Punishment for Aggression	21.5 (16)	34 (9)	7.5 (10)	4 (15)
Punishment for Dependency	19 (12)	22 (11)	10 (13)	8 (14)
Parent's Aggression	16 (13)	22 (11)	18 (11)	4 (15)
Lack of Social Participation	16 (19)	15.5 (6)	22.5 (6)	5 (19)
Shame at Home	12 (25)	18 (12)	18 (11)	2 (2)
Shame Out of Home	22 (14)	10.5 (12)	25 (10)	10 (14)

NOTE.—Number in parenthesis refers to frequency of subjects in cells (total $N = 50$). Median school aggression score for entire group of 50 children = 15.

ethnicity-aggression relationship. The high ethnic woman, married either to a low or high ethnic man, perhaps brings Old World values into her relationship with him, deferring to him as the decision-maker, authoritarian, etc. Their roles are well defined. On the other hand, the low ethnic woman, married to the high ethnic man, has a different conception of the marital relationship which then conflicts with what is expected by the husband, value conflicts ensue, role definitions are not agreed upon, and the children are deviant, at least as far as peer rating of aggression is concerned.

The importance of getting evaluations from both mother and father is again seen in the relation of confessing to aggression. It is only for those children whose mother and father agree that the child does not deny his guilt that the aggression score tends to be low. When both agree he does deny, the aggression score tends to be high. When there is disagreement between mother and father on this variable, the result is closer to the median for the entire group.

Parental disharmony operates differently. When both mother and father agree that parental disharmony is low, the child's aggression score tends to be low. When both agree parental disharmony is high or when father says it is high and mother says it is low, the child's aggression score tends to be around the median for the whole group. However, when the mother says disharmony is high and father says it is low, the child's aggression score tends to be very high. Thus, the mother's rating seems to be the crucial one here as it does in some aspects of the rejection-aggression relationship. But the father's role in the latter relationship is also important. Both mother and father must be high on rejection for the child's aggression score to be elevated, but only a low score for mother is needed for the child's aggression score to be low. When father is low and mother high, the aggression score is at the median for the whole group. In other words, there must be two rejecting parents to make for a high aggressive child, but only the mother's lack of rejection results in a low aggressive child. Quite the opposite is true for punishment for aggression. Here it is just when the father is high that the child's aggression score tends to be high. This is especially true when the father is high and the mother is low. When father is low in punishment for aggression, regardless of mother's score, the child tends to be low on aggression. With punishment for dependency, only low punishment by the father is associated with low aggression. When the father is a high punisher for dependency, regardless of whether mother is high or low, the child's aggression score tends to be moderately elevated.

The results with parent's aggression, when data from mother and father are considered jointly, indicate that an erroneous interpretation can be drawn if only one parent is taken into account. The father's importance as a role model for aggression was suggested in the results mentioned above; however, it is only when both mother and father rate themselves as low on aggressive behavior that the child's aggression score tends to be low. When either or both rates himself as high, the child's score tends to be at the median or slightly higher. The same is true for lack of social participation and shame at home; however, for shame out of home it is the low mother only who tends to have the low aggressive child. It is interesting that there are only two families in which both mother and father rate themselves as low on shame at home. In general, however, it should be apparent from this rudimentary analysis that, for a more complete picture of the effect of socialization practices and attitudes on child behavior, the contribution of both mother and father to this effect must be taken into account.

SUMMARY

Sixty subjects were selected out of a pool of 158 children on the basis of scores on an aggression sociometric administered in their classrooms. All of the mothers and 50 of the 58 available fathers of these children were then independently interviewed to gain information on a number of presumed psychological and social antecedents to aggressive behavior. It was found that mothers and fathers did not agree to an appreciable degree in rating either their children's behavior or their interactions with their children. Even on those variables where agreement between mother and father was moderately high, the relation to other variables was not the same for the two groups of parents. There was a consistent shrinkage in size of coefficient when proceeding from correlations between predictor and criterion which were both obtained from the same parent to correlations when the predictor was obtained from one and the criterion from the other. When an outside, independently obtained criterion was used, very often the father's scores related more adequately than did the mother's scores. The most meaningful and theoretically interesting relations were found when predicting from a combined mother and father score to the independent criterion. Only recently have studies emphasizing the importance of the father in socialization of the child begun to appear. These results are further evidence of his importance both as a new method and a new dimension in childrearing research.

REFERENCES

1. BANTA, T. J, WALDER, L. O., & ERON, L. D. Convergent and discriminant validation of a childrearing survey questionnaire. Paper read at East. Psychol. Ass., New York City, April, 1960.
2. BRONFENBRENNER, U. Family structure and development. Presidential address, Div. 7, Amer. Psychol. Ass., Washington, D.C., September, 1958.
3. CHAMPNEY, H. Measurement of parent behavior. *Child Develpm.,* 1941, 12, 131-166.
4. CURETON, E. E. Rank-biserial correlation. *Psychometrika,* 1956, 21, 287-290.
5. DAVIDSON, K. S., SARASON, S. B., LIGHTHALL, F. F., WAITE, R. R., & SARNOFF, I. Differences between mothers and fathers of low anxious and high anxious children. *Child Develpm.,* 1958, 29, 155-160.
6. GUILFORD, J. P. *Psychometric methods.* (2nd Ed.) McGraw-Hill, 1954.
7. LEVIN, H. Discussion of symposium: some methodological considerations in childrearing research. Paper read at East. Psychol. Ass., New York City, April, 1960.
8. LEVIN, H., & SEARS, R. R. Identification with parents as a determinant of doll play aggression. *Child Develpm.,* 1956, 27, 135-153.
9. MILLER, D. R., & SWANSON, G. E. *The changing American parent.* Wiley, 1958.
10. PETERSON, D. R., BECKER, W. C., HELMER, L. A., SHOEMAKER, D. J., & QUAY, H. C. Parental attitudes and child adjustment. *Child Develpm.,* 1959, 30, 119-130.
11. SEARS, R. R., MACCOBY, E. E., & LEVIN, H. *Patterns of child rearing.* Row, Peterson, 1957.
12. SEARS, R. R., PINTLER, M. H., & SEARS, P. S. Effects of father separation on preschool children's doll play aggression. *Child Develpm.,* 1946, 17, 219-243.
13. SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, P. S. Some childrearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.,* 1953, 47, 135-234.

Q-SORT DESCRIPTIONS OF FIVE-YEAR-OLD CHILDREN BY THEIR PARENTS[1,2]

GENE R. MEDINNUS*
San Jose State College

Although the plea for theory construction and new conceptualizations is a frequently repeated one in surveys of the parent-child literature, the need for the development of new methodological tools in child research is also recognized. The study reported here is an exploratory attempt to evaluate the use of the general Q methodology with parents of preschool children. The present paper will describe the procedures followed in developing and administering the item pools and the general results obtained with regard to the reliability of the technique and interparent agreement in relation to such first order variables as sex of child and sex of parent. Subsequent reports in this series will include data concerning the relation between interparent agreement and the child's adjustment as well as findings in the area of the consistency of interparent agreement from one measuring instrument to another.

An attempt has been made to overcome some of the shortcomings present in many parent-child researches: nonurban parents from the lower and

---

* Psychology Department, San Jose State College, San Jose 14, California.

[1] This research was completed under the sponsorship of the Elizabeth McCormick Memorial Fund of Chicago.

[2] A summary of the present paper was presented at the West. Psychol. Ass., San Jose, California, April, 1960.

Child Develpm., 1961, 32, 473-489.

lower-middle classes participated as subjects, and information was obtained from the fathers as well as from the mothers. Several writers (1, 11) have pointed out the meager amount of parent-child research data which has been obtained from fathers. Sigel (11) stressed the incongruity of this situation in view of the prominent role which the father plays in Freudian psychology. In an analysis of some of the parent-child literature using a role theory framework, Brim (1) noted that mothers outnumbered fathers two to one as sources of information; this he attributed to the greater availability of mothers as research subjects.

In general, the Q-sort technique, described in detail by Stephenson (12), involves a task which requires the subject to sort a large number of descriptive phrases into groups or piles which he feels characterize himself to a greater or lesser degree. Typically the item pool contains over a hundred items which are sorted into 11 to 13 groups ranging from those which the subject feels are "most characteristic of myself" to the group of items at the other end of the continuum labeled "least characteristic of myself." The number of items to be placed in each pile is usually designated in advance. When used in this manner, the method allows for a systematization of an individual's self-perception as well as for comparisons among individuals.

Rogers (9) has used the Q-sort technique to obtain a picture of the client's self-perception before and after therapy, as well as of the client's view of the "ideal self." The method has been used also to compare interviewers' or clinicians' judgments (descriptions) of clients. Such sorters commonly observe that they have insufficient information about the client to make judgments concerning every item. In this respect it would seem that the technique is especially appropriate for use by parents in describing their children since the parent certainly is most intimately acquainted with the characteristics and behavior of his child and is the best qualified informant in this regard.

PROCEDURE

Subjects

The 76 parents, 19 sets of parents of boys and 19 sets of parents of girls, in the present study were all parents of 5-year-olds who were subjects in an investigation of school readiness. The location of the field study was a small midwestern community of approximately 5,000 residents. The 5-year-old sample itself was not completely representative of the entire population of 5-year-olds in the community so far as intelligence and father's occupation are concerned. The mean IQ of the 5-year-olds was 111.6 according to the Stanford-Binet (Form L) intelligence tests administered by the writer. The socioeconomic status of the children's families was approximately evenly distributed between the upper-lower and lower-middle classes according to Warner's Index of Status Characteristics (13). The mean ages of the mothers and fathers were 32.1 and 34.1 years, respectively, and the corresponding mean years of education were 12.0 and 12.2.

Construction of the Item Pool

In assembling the pools of items to be used, an attempt was made to obtain as complete a coverage as possible of characteristics which pertinently and significantly describe the 5-year-old level. Child development and child psychology texts and literature were consulted, and numerous child behavior rating scales were examined. Overlapping descriptions of behavior were eliminated, and those items were used which showed the greatest frequency of mention in the above sources as well as those thought by clinicians to be most predictive of later adjustment (research evidence on this point is meager, however). In addition, since the larger study dealt with first grade adjustment, several items were taken from a first grade adjustment scale constructed by the writer (8), based on interviews with 25 first grade teachers. In general, characteristics covering the following major areas of child behavior were included: emotional behavior, social behavior, physical status and motor behavior, and intellectual abilities. So far as possible, the items were phrased colloquially in language familiar and comprehensible to parents.

In the preliminary pilot investigation with a small group of parents, positive and negative characteristics were included in a single pool. The parents were asked to sort the items into seven piles, ranging from "most characteristic of my child" to "least characteristic of my child." Bimodal distribution of items resulted in all cases; a majority of the positive items were placed near the "most characteristic" end of the continuum, while most of the negative items fell at the "least characteristic" end. This resulted in little discrimination among the items for the various parent sorters. To obtain such discrimination it was necessary to divide the positive and negative characteristics into separate pools. Additional evidence for the necessity of controlling the social desirability of items is presented in a study by Edwards (3) in which 50 male and 50 female college students made "self-sorts" using 135 items which had been rated previously on a scale of social desirability. Product-moment correlations of .84 and .87 were found between the mean weights assigned the items by the sorters and the measure of social desirability of the items for the male and female students, respectively.

It was decided initially to reduce the number of items customarily used in the Q-sort procedure. One purpose of this was to make the task more acceptable to the parents. Also, research workers using the Q-sort technique often criticize existing item pools on the grounds that they include both phenotypic and genotypic items. Comparisons and judgments must be made between readily observable characteristics and those which involve inferentially-drawn interpretations. Following is an example of each of the two types of items: phenotypic, "Is of average height"; genotypic, "Has an unresolved Oedipus complex." Therefore, the present items were limited to those describing fairly overt types of behavior and/or to those about which parents could make judgments based on day-to-day observation and contact with their children. The number of items in each of the two pools (positive and negative) was limited to 42.

Unforced distributions (the number of items to be placed in each pile not designated) were used in the pilot investigation, but again little item discrimination resulted, since in the sort of the positive pool of items a majority of them were placed at the "most characteristic" end of the distribution while the reverse was true for the negative pool of items. Since it was apparent that a forced distribution was advisable, only the shape of the distribution remained to be determined. The parents in the pilot study were asked to sort the items using both a rectangular and a seminormal distribution. They reported much difficulty in the sort requiring a seminormal distribution of the items; this was due in large part to the very small number of items at either end of the distribution. The parents felt also that the large number of items placed in the middle pile added little to the description of their particular child. This last point is supported by the findings of Livson and Nichols (6), who concluded that the forced rectangular distribution requires the maximum possible number of interitem discriminations. They also present evidence to indicate that "Making more discriminations tends to increase the reliability of the total information carried by the sort" (p. 163).

To obtain a measure of the parents' value system with regard to child behavior and attributes, they were requested also to sort the two pools of items in terms of the "ideal 5-year-old." This is fairly straightforward for the positive pool of items; the items were sorted on a continuum from "most important for the ideal 5-year-old to possess" to "least important for the ideal 5-year-old to possess." Since one would not want an "ideal" 5-year-old to possess any of the characteristics included in the negative pool, the procedure here involved arranging the items on a continuum from "least bad for the ideal 5-year-old to possess" to "worst for the ideal 5-year-old to possess."

In summary, the technique involved two sets of items, one containing only positive traits or characteristics and the other only negative aspects of behavior. Each pool consisted of 42 items typed on separate cards. The cards were sorted into seven piles ranging from "most characteristic of my child" to "least characteristic of my child," with six cards in each pile. Following is a list of the 84 items.

PLUS POOL	MINUS POOL
1. Can take disappointments calmly	1. Can't seem to stick at anything
2. Is able to give and accept affection from others	2. Gets mad easily

PLUS POOL (*continued*)	MINUS POOL (*continued*)
3. Is obedient; usually does what he is told to do	3. Is easily influenced by other children; can be talked into doing things
4. Girl: very aware of herself as being a girl Boy: very aware of himself as being a boy	4. Always thinks he has to be the best when he's doing things with other children
5. Gets over being mad quickly	5. Has temper tantrums
6. Has a pretty even disposition—doesn't have many ups-and-downs in how he feels	6. Is nervous and highstrung
7. Is interested in learning new things; asks many questions	7. Doesn't make friends easily—takes quite awhile to get acquainted with children he doesn't know
8. Is willing to take care of his own possessions, such as toys	8. Is a daydreamer
9. Has a lot of energy and pep; doesn't get tired very quickly	9. Likes to be the center of the stage; is a show-off
10. Gets along fine in new (or strange) situations	10. Takes his hurts seriously; becomes worried about a slight bruise
11. Catches on to new ideas quickly—for example, is quick at learning how to play new games	11. Usually takes a back seat when playing with other children
12. Likes to be read to	12. Lacks drive; no spark
13. Is able to take correction and criticism without getting all upset	13. Gives up easily without really trying
14. Can be counted on to carry out a job which he is given to do	14. Is very babyish when sick or hurt; likes to get a lot of sympathy from others
15. Is forgiving; doesn't hold grudges	15. Seems to be afraid of quite a few things
16. Is able to keep himself busy; can figure out for himself what to play or what to do next	16. Every little hit from another child calls forth a full battle
17. Has a good imagination	17. Is pretty stubborn at times
18. Uses good language and words for his age; has a good vocabulary	18. Weak physically; poor health
19. Is in good physical condition; usually healthy	19. Gets real mad and upset if he can't have his own way
20. Gets along well with other children; is well liked by other children	20. Often blames others for things he does
21. Is sympathetic toward others' hurts and misfortunes	21. Whines or pouts quite a bit
22. Is a self-starter; has a lot of initiative	22. Cries easily

Plus Pool (*continued*)	Minus Pool (*continued*)
23. Has a good sense of humor; can understand jokes	23. Doesn't seem to be able to plan ahead very well
24. Is cooperative in family affairs	24. Is easily confused in carrying out a simple task; forgets what to do next
25. Is able to follow directions; has a good memory for what he is told to do	25. Has trouble holding onto small objects, such as scissors, jigsaw puzzle pieces, etc.
26. Is good looking, an attractive child	26. Is not very considerate of the property rights of others
27. Is considerate of other people	27. Doesn't like to play with other children
28. Is fair—plays fair with other children	28. Is unpredictable; I never know what he's going to do next
29. Has self-confidence	29. Is always on the go; restless
30. Is usually happy, cheerful	30. Has a habit of lying about some things
31. Sticks up for himself; doesn't let others run over him	31. Feelings are easily hurt
32. Is willing to confide in others (parents, teacher, etc.)	32. Doesn't show common sense sometimes
33. Is usually a leader in the group he plays with	33. Immature—doesn't act his age
34. Is willing to share his things with others	34. Is afraid to try new things, such as new games, new play equipment
35. Has a lot of courage; is brave	35. Sulks and mopes around for quite awhile after being punished
36. Is wide awake as far as what is going on around him	36. Doesn't get along too well in a group of children
37. Has a good idea of what he is and isn't capable of doing; doesn't think he can do more than he really can	37. Is impulsive; acts on the spur of the moment
38. Shows good judgment; is able to profit from experience	38. Is noisy and loud
39. Is courteous and mannerly	39. Is shy and bashful around strangers
40. Likes to explore and investigate things	40. Is boastful or bragging at times
41. Knows the difference between right and wrong	41. Is sort of a slow mover—takes his time
42. Seldom picks fights with other children; is not quarrelsome	42. Is a worrier

Method

Individual evening appointments at the convenience of the parents were made in advance; in every case the investigator visited the home for the data collection. The sorting was done independently and simultaneously by the mother and the father, seated at opposite ends of a table. The directions for sorting were given, and any questions were answered. After each sort was completed, the items in each pile were placed in appropriately marked envelopes. The order of presentation of the item pools was as follows: the positive and negative pools, in that order, were sorted to describe the parents' own 5-year-old (real sort); then the two pools were sorted to describe the ideal 5-year-old (ideal sort). The items in each of the two pools were numbered from 1 to 42. After each set of parents completed the sorting, their placement of the items was recorded, and the items were rearranged in numerical order before the task was presented to the next set of parents. This procedure standardized the order placement of the items so as to eliminate possible effects of differences in interitem context from parent to parent.

Reliability of the sorts, with regard to their stability over time as determined by the test-retest method, was also obtained. Ten sets of parents, five of them parents of girls and five of boys, selected at random, resorted the pools of items four months, plus or minus one week, after their initial sorts.

Results

Feasibility of the Technique

The parent subjects appeared to understand readily the procedure of Q-sorting although none had had previous acquaintance with the technique. Likewise, the phrasing and meaning of the items were comprehended for the most part. The task was approached with interest and performed conscientiously by the parents. Some of the subjects experienced difficulty in deciding upon the placement of certain items, but, in general, the task produced no negative reactions. In fact, many of the parents enjoyed the sorting and several observed that they had never formalized their perceptions of their child before. It can be said, therefore, that the Q-sort technique is a feasible one for the expression of judgments by parents with regard to their children.

Although the exact time required for the sorting of the four item pools was not recorded, it varied from approximately 45 to 90 minutes.

Reliability

The sort-resort reliability figures are tabulated in Appendix B.[3] The reliability coefficients were obtained in the following manner: A correlation coefficient was computed between each parent's original sort and his subsequent sort for each of the four sortings (plus and minus pools for the real and ideal sorts). The means of reliability coefficients for the four sortings by the 10 sets of parents were calculated. The reliability coefficients ranged from .01 to .83; the means for the real sort, plus and minus pools, and for the ideal sort, plus and minus pools, were .57, .55, .51, and .53, respectively.

---

[3] The appendixes (A, B, C) have been deposited as Document number 6719 with the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C. A copy may be secured by citing the Document number and by remitting $2.50 for photoprints, or $1.75 for 35 mm. microfilm. Advance payment is required. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress.

Interparent Coefficients for Real and Ideal Sorts

The correlation coefficients for interparent agreement in their descriptions of their child (real sort) and for the ideal 5-year-old (ideal sort) appear in Appendix C. The mean coefficients for the total group of fathers and mothers, as well as for the parents of girls and the parents of boys, are summarized in the first four columns of Table 1. Concerning these coefficients, it can be noted that (a) the mean interparent correlations are higher for the real sorts than for the ideal sorts, indicating that parents agree more in how they perceive their child than in their expectations and goals for him; and (b) the mean correlations for the parents of boys are higher than for the parents of girls.

TABLE 1

MEAN CORRELATION COEFFICIENTS FOR INTERPARENT AND INTRAPARENT AGREEMENT

	INTERPARENT AGREEMENT				INTRAPARENT AGREEMENT			
	Real Sort		Ideal Sort		Plus Pool		Minus Pool	
Parents	*Plus Pool*	*Minus Pool*	*Plus Pool*	*Minus Pool*	*Mother*	*Father*	*Mother*	*Father*
Of Girls	.37	.37			.25	.19	.40	.20
Of Boys	.43	.49			.28	.34	.33	.32
Combined ...	.40	.43	.28	.30	.27	.26	.36	.26

Intraparent Coefficients for Real and Ideal Sorts

An individual parent's two sorts (real and ideal) of the two pools of items were correlated. These coefficients would seem to be an index of the degree to which the parent's perception of his child agrees with his description of the ideal 5-year-old and might be interpreted as the extent of parental acceptance of the child. The coefficients are listed in Appendix C and their means are summarized in the last four columns of Table 1. The means are generally low. They are higher for the fathers of boys than for the fathers of girls on both the plus and minus pools. On the minus pool the mean correlations for the mothers of girls are higher than for the mothers of boys, but the trend seems to be in the opposite direction on the plus pool.

Also, for the girls, the mean correlations of the mothers are higher than those of the fathers, but there is no such difference between the parents of boys.

Rank Order and Mean Placement of the Items

The mean placement of each item in the two pools by the 76 parents is given in Appendix A. The mean placement of the items on the real sort for the various subgroups of parents, as well as for the fathers and mothers separately on the ideal sort, is also listed in Appendix A.

TABLE 2

ITEMS DIFFERENTIATING BETWEEN MOTHERS AND FATHERS OF BOYS ON REAL SORT

Item	MEAN PLACEMENT *Mothers*	*Fathers*	*p*
Plus Pool			
4. Very aware of himself as being a boy	4.3	2.8	$< .05$
5. Gets over being mad quickly	3.1	4.0	$< .05$
8. Is willing to take care of his own possessions, such as toys	1.9	2.6	$< .05$
Minus Pool			
10. Takes his hurts seriously; becomes worried about a slight bruise	2.0	3.6	$< .01$
13. Gives up easily without really trying	3.5	2.2	$< .01$
33. Immature—doesn't act his age	2.2	1.4	$< .05$
39. Is shy and bashful around strangers	2.6	3.8	$< .01$

NOTE.—To interpret the means in this and the following tables, it is necessary to point out that the piles into which the items were sorted were numbered from "6" (most characteristic of my child) to "0" (least characteristic of my child) on the real sort. For the ideal sorts pile "6" denoted the most desirable end of the continuum while pile "0" was the least desirable end.

The mothers' and fathers' placements of each item in the ideal sorts were summed separately, and rho was computed between the two groups' rank ordering of the items. Rank order coefficients of .84 and .93 were obtained for the plus pool and minus pool, respectively. The significance of the difference in the mean placement of the items by the following parent groups was calculated: the mothers versus the fathers of boys, the mothers versus the fathers of girls, the fathers of boys versus the fathers of girls, and the mothers of girls versus the mothers of boys. The t test was used to assess the significance of the difference between the various means. Table 2 gives those items whose mean placement for the boys were significantly different between the fathers and mothers on the real sort. Comparable items for the girls are tabulated in Table 3. Tables 4 and 5 list those items whose mean placement was significantly different for the fathers of boys versus the fathers of girls and for the mothers of boys versus the mothers of girls, respectively.

Those items whose mean placement was significantly different between the fathers and mothers on the ideal sort are given in Table 6.

TABLE 3

ITEMS DIFFERENTIATING BETWEEN MOTHERS AND FATHERS OF GIRLS ON REAL SORT

Item	MEAN PLACEMENT *Mothers*	MEAN PLACEMENT *Fathers*	*p*
Plus Pool			
11. Catches on to new ideas quickly—for example, is quick at learning how to play new games	3.4	4.4	$< .05$
23. Has a good sense of humor; can understand jokes	3.2	2.0	$< .05$
Minus Pool			
3. Is easily influenced by other children; can be talked into doing things	4.5	3.6	$< .05$
7. Doesn't make friends easily—takes quite awhile to get acquainted with children he doesn't know	2.8	1.3	$< .01$
25. Has trouble holding onto small objects, such as scissors, jigsaw puzzle pieces, etc.	.6	1.8	$< .05$
38. Is noisy and loud	3.1	4.2	$< .05$

TABLE 4

ITEMS DIFFERENTIATING BETWEEN FATHERS OF BOYS AND FATHERS OF GIRLS ON REAL SORT

Item	MEAN PLACEMENT *Fathers of Boys*	MEAN PLACEMENT *Fathers of Girls*	*p*
Plus Pool			
9. Has a lot of energy and pep; doesn't get tired very quickly	4.9	3.2	$< .01$
21. Is sympathetic toward others' hurts and misfortunes	2.2	3.5	$< .05$
Minus Pool			
3. Is easily influenced by other children; can be talked into doing things	5.0	3.6	$< .01$
12. Lacks drive; no spark	.8	2.1	$< .01$
25. Has trouble holding onto small objects, such as scissors, jigsaw puzzle pieces, etc.	.7	1.8	$< .05$
27. Doesn't like to play with other children	1.0	2.2	$< .05$
37. Is impulsive; acts on the spur of the moment	5.0	3.6	$< .01$

TABLE 5

ITEMS DIFFERENTIATING BETWEEN MOTHERS OF BOYS AND MOTHERS OF GIRLS ON REAL SORT

Item	Mean Placement		
	Mothers of Boys	*Mothers of Girls*	*p*
Plus Pool			
1. Can take disappointments calmly	1.2	3.0	$< .01$
17. Has a good imagination	4.8	3.4	$< .05$
40. Likes to explore and investigate things	3.5	1.9	$< .05$
Minus Pool			
2. Gets mad easily	4.5	3.1	$< .01$
11. Usually takes a back seat when playing with other children	2.1	3.5	$< .05$
16. Every little hit from another child calls forth a full battle	3.4	2.1	$< .05$
38. Is noisy and loud	4.2	3.1	$< .05$
39. Is shy and bashful around strangers	2.6	4.3	$< .05$

TABLE 6

ITEMS DIFFERENTIATING BETWEEN MOTHERS AND FATHERS ON IDEAL SORT

Item	Mean Placement		
	Mothers	*Fathers*	*p*
Plus Pool			
10. Gets along fine in new (or strange) situations	3.3	2.4	$< .05$
35. Has a lot of courage; is brave	.8	1.6	$< .05$
Minus Pool			
21. Whines or pouts quite a bit	1.6	2.3	$< .05$
22. Cries easily	2.2	3.1	$< .05$
37. Is impulsive; acts on the spur of the moment	5.1	4.2	$< .05$
38. Is noisy and loud	4.2	3.3	$< .05$

DISCUSSION

Because the Q-sort technique possesses a number of advantages over other methods of obtaining information concerning parents' perception of and attitudes toward their children, its use is indicated in many situations where such information is considered necessary or desirable. The technique is superior certainly to the method of parental rating of the child on various scales or variables. It has been the writer's experience that with such a method the majority of parents tend to rate their child "average" or above, which, of course, eliminates variation among parents as well as variation from one variable to another for any one parent rater. Consequently, little significant information is gained from this approach. Check lists typically provide even less information. Advantages of the Q-sort technique include: (a) the derivation of an index of agreement between the parents either through correlational methods or through simply a count of the difference in placement of each of the items employed; (b) the derivation of an index of agreement between a single parent's real and ideal sorts; (c) in effect, the placement of each item can be considered as the parent's rating of the child on that variable which also provides comparisons among the various items or characteristics; (d) an unsubstantiated observation by the investigator would indicate that the procedure tends to result in increased communication between parents, especially with regard to the child under consideration. Animated discussions between parents frequently followed the completion of the sorting. The implications of this for clinical approaches to parent counseling are numerous.

While the reliability coefficients are not high, several factors might have tended to reduce them. The time interval of four months, combined with the fluctuating nature of child behavior at this age level, was perhaps an important factor in this regard. A time interval of one to three weeks between sorts probably would have raised the reliability coefficients considerably. That the changeableness of child behavior at the 5-year level was a factor influencing the magnitude of the reliability coefficients was substantiated by the observations of several parents during the resorting when they noted the extent to which their child had changed with regard to specific items of behavior in the months intervening between the two sorts.

The moderate reliability coefficients strongly suggest also that single measures of parent attitudes and of parental perceptions of their children are not adequate. This point has important implications for those investigations which attempt to relate parent attitudes to child behavior and adjustment. Few, if any, studies exist in the research literature in which concern is shown for variations over time in the parent variable under consideration. The writer has stressed elsewhere (7) the necessity for recognizing the possibility of little long-term constancy in various parent attitudinal characteristics; a number of findings were mentioned which indicate the existence of shifts in parent attitudes, parent behavior, and the like. It might be postulated that the nature, extent, and quality of these shifts are as important determiners of child adjustment as are measures of parent attitudes elicited at any one point in time.

Interestingly, the reliability coefficients were no higher for the ideal sorts than for the real sorts. This seems contrary to expectation since one might define the ideal sorts as representing the parents' set of values regarding child behavior as well as the goals which the parents hope to attain through their child training procedures. Both of these would seem to be fairly stable over time. It might be argued, however, that those specific aspects of the child's behavior which are sources of difficulty to the parent at any given time are those which the parent is most likely to consider important. For example, if a child has been ill a great deal during a certain period, the parent might attach great importance to item 19, "Is in good physical condition; usually healthy," when describing the ideal child. This is to say that the day-to-day changes in a child's behavior and the problems he presents influence the parent's description of the ideal child as well as his description of the child himself.

That the interparent agreement coefficients were higher for the real than for the ideal sorts is understandable perhaps in view of the objective referent for the former sort as compared with the latter. It might be postulated also that there is greater interparent communication concerning the actual child, his behavior, and his characteristics than there is in the area of parental expectations, demands, and goals of childrearing.

The trend toward higher interparent correlations for parents of boys than for parents of girls is difficult to interpret. Perhaps boys, by the very nature of their high activity level and personality traits in general, necessitate greater interparent discussion concerning their behavior. With little research on this point to turn to for an explanation, one might assume that the behavior of boys is more obvious, more overt, and less subtle than is true for girls; consequently, this would lead to greater interparent agreement in perception of boys than of girls. In any case, further verification of the present finding would require research designed to sort out the possible causes for this.

Again, with regard to the intraparent correlation coefficients, there are few data in the literature to assist in interpreting the trends found. One investigation (4), using a paper-and-pencil test designed to measure parental acceptance of children, found that mothers scored higher (i.e., greater acceptance) than did fathers. This would lead us to predict higher intraparent coefficients (real-ideal correlations) for mothers than for fathers; such was not found to be the case in the present study, however. Further analysis (2) of the same data revealed no differences between mothers and fathers with regard to acceptance of children of one sex or the other.

Several findings, on the other hand, would lead us to predict higher real-ideal coefficients for mothers of girls than for mothers of boys. In a study of maternal attitudes and practices in child rearing from birth to the age of 2 years, Hubert and Britton (5) found that mothers seemed to enjoy their daughters more than mothers did their sons. They conclude that "Mutual identification between mother and daughter and between father and son is stronger than mother-son and father-daughter identification, at least as seen indirectly by the mothers" (p. 219). In another study in which the data were obtained again through interviews with the mothers, Sears, Maccoby, and Levin (10) found that in infancy girls were somewhat more warmly treated than were boys. While both of these studies refer specifically to the period of infancy, it is interesting to note that in the present study, data regarding mothers' perceptions of their 5-year-olds do not indicate that mothers accept their daughters more than do mothers accept their sons—that is, if the real-ideal correlation coefficients can be interpreted as indexes of parental acceptance. The coefficients listed in the last four columns of Table 2 would indicate, however, a trend toward higher intraparent coefficients for fathers of boys than for fathers of girls. Also, again on the basis of the present interpretation of the real-ideal correlation coefficients, it would appear that mothers accept their daughters more than do the fathers. These latter two trends are in agreement with the conclusion reached by Hubert and Britton concerning stronger same-sex than opposite-sex parent-child identification.

The rank-order coefficients obtained between the total group of mothers' as compared with the fathers' placement of the items in the ideal sort indicate a high degree of agreement between mothers and fathers concerning the value they place on general areas of behavior as well as on specific items of behavior. Low coefficients resulted, however, from correlating the ideal sorts of the individual pairs of parents. These findings would suggest that, while in general there is little parent sex difference with regard to their evaluation of the importance of various aspects of child behavior, any one set of parents may show a great amount of disagreement. There were some indications also that, as groups, fathers and mothers agreed more on the significance attached to undesirable aspects of behavior than to the more favorable aspects.

With regard to those items whose mean placement on the real sorts was significantly different depending upon the sex of the parent and the sex of the child, it is apparent that in general their number is little greater than that expected by chance. Again, as in the comparison of the mean placement of the items in the ideal sorts by the groups of fathers and mothers, there would seem to be no consistent differences between fathers and mothers in their perception of their children regardless of the sex of the child. So far as the sex of the child is concerned, these data, though dealing with a different aspect of the mother-child relationship, agree with the Sears conclusion, based on interviews with a large sample of mothers concerning various childrearing practices, that "There were surprisingly few dimensions on which the sexes were differently treated" (10, p. 401).

The finding that there is fairly close agreement between groups of fathers and groups of mothers in their perception of their children (real sorts) and in the importance which they attach to various items of child behavior (ideal sorts), but only moderate agreement between individual pairs of parents, would seem to have important implications for role theory in the area of parent-child relations. It is not necessary to be concerned with differences between mothers in general and fathers in general in their role prescriptions for the child since such differences are small. More importantly, it is necessary to describe the differences existing for a specific set of parents; it is this difference which would affect most crucially the child's behavior and adjustment (assuming that there is a relation between these variables).

Although numerous discussions of role theory have included comments concerning the role conflicts produced by varying role demands and role prescriptions which an individual must satisfy, there are few research data to support this notion. Specifically, within the role theory framework, it would be predicted that great discrepancy between the father and the mother in their role prescriptions for child behavior would lead to conflict and subsequent maladjustment on the part of the child. It might be postulated, on the other hand, that a certain amount of discrepancy in role demands is useful insofar as it provides the child with practice and experience in role playing. The corollary of this would be the hypothesis that there is a curvilinear relation between the amount of discrepancy between the parents in their role prescriptions for the child and the child's consequent conflict and maladjustment. With too little practice in role playing in the home, the child might well experience difficulty outside of the home when required to meet the demands of other significant adults. When there is too large a discrepancy between the parents in their role prescriptions, conflict is produced in the child in two ways: first, the child is unable to meet the role prescriptions of both parents, especially when the prescriptions demand opposing types of behaviors; second, the child faces the problem of determining exactly when each of the two roles is required of him.

In a role theory context, Brim (1) pointed out the lack of specificity of the research data with regard to the sex of the parent and the sex of the child. The present findings would suggest that this is not so important as concern with role prescriptions and performances specific to any given family. It is the writer's contention, then, that in general few differences exist in attitudes and behavior as a function of the sex of the parent and the sex of the child. Research in the area of parent-child relations needs to be concerned, rather, with the differences which obtain for one set of parents and in relation to their offspring and in relation to their specific family constellation. Idiographic rather than nomothetic data seem to be indicated.

SUMMARY

The purpose of the present study was to evaluate the use of the general Q methodology with parents of preschool children and to explore the extent of interparent Q-sort agreement in the area of parental perception of child behavior.

The sample consisted of 76 parents, 19 sets of parents of 5-year-old boys and 19 sets of parents of 5-year-old girls. The mean IQ of the children was 111.6. The socioeconomic status of the families was divided approximately

equally between the upper-lower and lower-middle classes according to Warner's Index of Status Characteristics. The mean ages of the fathers and mothers were 32.1 and 34.1, and mean years of education 12.0 and 12.2, respectively.

Both parents of each set sorted, independently, two pools of 42 items each descriptive of the behavior of 5-year-olds. One pool contained only positive items (plus pool) and one contained negative items only (minus pool). The items were sorted into seven piles, with six items per pile, ranging from "most characteristic of my child" to "least characteristic of my child." In addition to this "real" sort, the items were also sorted by the parents to describe the "ideal" 5-year-old. Product-moment correlation coefficients were computed between each set of parents on the two sorts (interparent agreement) and between each parent's two sorts (intraparent agreement). Reliability figures were obtained from 10 sets of parents who resorted the items with a four-month interval between sorts.

The results and conclusions may be summarized as follows: (a) The mean interparent correlations were higher for the real sorts than for the ideal sorts. This would indicate that parents agree more in how they perceive their child than in their behavioral goals (role prescriptions) for him. (b) If the intraparent coefficients (real-ideal sorts) can be interpreted as the degree to which the parent accepts the child's behavior, it appears that there is only moderate acceptance of the child's behavior among the present sample of parents and that there is wide variability among parents in this regard. (c) Although the individual sets of parents did not agree closely in their ideal sorts, the mean placement of the items by the total group of mothers versus fathers showed high agreement. (d) Few significant differences in the mean placement of items were found between various subgroups of parents (e.g., fathers of girls versus fathers of boys) in their real sorts. That there were few differences as a function of the sex of the parent and the sex of the child was discussed in relation to role theory. The necessity for an idiographic approach in the parent-child area was suggested.

REFERENCES

1. BRIM, O. G., JR. The parent-child relation as a social system: I. Parent and child roles. *Child Develpm.*, 1957, 28, 343-364.
2. BURCHINAL, L. G. Mothers' and fathers' differences in parental acceptance of children for controlled comparisons based on parental and family characteristics. *J. genet. Psychol.*, 1958, 92, 103-110.
3. EDWARDS, A. I. Social desirability and Q-sorts. *J. consult. Psychol.*, 1955, 19, 462.
4. HAWKES, G. R., BURCHINAL, L. G., GARDNER, B., & PORTER, B. M. Parents' acceptance of their children. *J. Home Econ.*, 1956, 48, 195-200.

5. Hubert, M. A. G., & Britton, J. H. Attitudes and practices of mothers rearing their children from birth to the age of two years. *J. Home Econ.,* 1957, 49, 208-223.

6. Livson, N. H., & Nichols, T. F. Discrimination and reliability in Q sort personality descriptions. *J. abnorm. soc. Psychol.,* 1956, 52, 159-165.

7. Medinnus, G. R. Research implications of several parent-child concepts. *Marriage Fam. Living,* 1959, 21, 329-333.

8. Medinnus, G. R. The development of a first grade adjustment scale. *J. exp. Educ.,* in press.

9. Rogers, C. *Client-centered therapy.* Houghton Mifflin, 1951.

10. Sears, R. R., Maccoby, E. E., & Levin, H. *Patterns of child rearing.* Row, Peterson, 1957.

11. Sigel, I. E. The need for conceptualization in research on child development. *Child Develpm.,* 1956, 27, 241-252.

12. Stephenson, W. *The study of behavior.* Univer. of Chicago Press, 1935.

13. Warner, W. L., Meeker, M., & Eels, K. *Social class in America.* Science Research Associates, 1949.

THE RELIABILITY OF MOTHERS' HISTORIES

CHARLES WENAR*

University of Pennsylvania School of Medicine

In view of the extensive use made of mothers' developmental histories of their children, surprisingly little is known concerning their reliability. Clinicians and experimentalists are aware of the fact that reconstructive data distort reality, but there are few studies of the nature and extent of this distortion. Certainly Macfarlane's findings are sobering: In the Berkeley longitudinal study she found that, by the time the infant was only 21 months old, ". . . the retrospective account (of the physical condition during pregnancy) was so unreliable that we have had to disregard it. . . . Weight at birth was reliably reported. The use of instruments was unreliably reported —only two-thirds of the mothers delivered with instruments reported this fact. The duration of labor showed an average discrepancy of three to five hours. . . . Illnesses, unless outstanding, were frequently forgotten" (6, p. 127). She concluded, "Where retrospective interview data are the only type available, the above findings should limit the over-optimistic use of them as factual" (6, p. 29).

---

* Department of Psychiatry, Hospital of the University of Pennsylvania, 34th and Spruce Streets, Philadelphia 4.

Child Develpm., 1961, 32, 491-500.

Since the interview shows every sign of remaining a popular research instrument and since many data concerning child development and childrearing practices are derived from mothers' reports, a methodological enquiry is appropriate. The purpose of this paper is to present a conceptual framework in which to study the validity and reliability of mothers' histories. As will be seen, the spadework for many of the methodological studies has been done in other contexts, and the studies themselves will reveal a good deal not only about reconstructive material but about the parent-child relationship as well.

The schema can best be introduced by fantasying what the Ideal Research Mother would be like. Such a mother could (a) faithfully record events, feelings, and interpersonal relations as they happen; (b) remember with perfect fidelity; and (c) have the power of total recall upon request. What the offspring of the IRM would be like need not concern us here, since she was created solely to illustrate three essential aspects of studying historical material. The first raises the question of the validity of the mother as an observer, while the second and third are concerned with her reliability or consistency over time.

VALIDITY OF INITIAL OBSERVATION

As a rule, a mother's history can be no more accurate than her original observations and evaluations of events, feelings, and interactions. (An exception to this rule will be discussed in the next section.) Yet, very little is known about this crucial and fascinating area. Arguing from certain clinical and experimental data, one could say the mother is the ideal observer since she is the person most involved and in most frequent and intimate contact with her child. Other kinds of clinical observations and research data could be cited to show that, by her very involvement, a mother is the person most vulnerable to distortions, denials, and the sundry tricks affect plays on cognition.

In its simplest form, determining the mother's validity is a matter of correlating her report with an independent measure of the event. For some kinds of factual information, criteria already exist, and there is reason to be optimistic about more becoming available in the near future. Recent advances in measuring infant responses and the increasing number of normative studies will aid appreciably in solving the criterion problem. For instance, if a mother says her infant is "very restless," this statement can be compared with measures of the infant's gross motor activity and normative data on a large group of infants. Other norms can be used to determine how accurate a mother is in her statements concerning the developmental status of her child or the timing and nature of her childrearing practices.

Attitudinal and evaluative statements are more difficult to validate. Unfortunately, such statements are heavily weighted in many clinical situations and research projects. When a mother says, "I don't pay any attention to his lisping," or "He never seems to want to be loved," her report may be extremely important in studying the etiology of speech defects or withdrawal tendencies, while being extremely difficult to validate.

Even in this area, however, there are encouraging signs. Smith (9) observed manifestations of dependency in the mother-child interaction and later interviewed the mothers concerning their techniques for handling dependency. She found that 70 per cent of the mothers reported using techniques they were observed to use and that ratings on techniques from the interview were positively related to the total observed techniques at the .05 level. She further found that mothers reported more accurately their use of negative direction than they did their use of positive direction. This was interpreted as meaning they were more conscious of actions which they considered bad. In other observational studies, it would be comparatively simple to ask the mother, at the end of the experiment, to rate the same or similar variables used by the judges. In this way, data can be accumulated on the kinds of behavior mothers can evaluate with a high degree of accuracy and the amount of error which is to be expected.

In such studies, mothers' evaluations are compared with those of judges. Although the latter are not as scientifically pleasing as the reading of a meter, the assumption is that the expert can rate the behavior and attitudes more objectively than the mother. A thoroughgoing skeptic might point out that experts can be victimized by their feelings or their social class or their hypotheses just as the mother can be blinded by her neurosis. Until the day a purely objective science of child development is devised, however, the field can do no better than rely on its experts, standardizing them and controlling them to the best of its ability.

As was mentioned above, such methodological studies can also advance our understanding of the mother-child relationship. Two examples will be briefly described, one concerning "invalid," the other concerning "valid" mothers.

Mothers who are poor observers, instead of being dismissed as sources of error, comprise a particularly interesting subgroup. In clinical settings, a mother's distortions are especially valued as furnishing keys to understanding her personality, her past experience, her unconscious attitudes toward the child and, consequently, her faulty childrearing practices. When a mother calls her child a "terrible liar," but can cite only a few instances in which incomplete homework was claimed to be done, this distortion illuminates both the nature of her value system and the emotional impact of her morality training on her child. Further research could proceed in a number of directions. One is to collect data on the mother's personality development in order to test the clinical hypothesis that present distortions arise from her own childhood conflicts in the same area. Thus, one could hypothesize that the mother who unrealistically complains that her infant is helpless and demanding has herself been inadequately loved as a child and thus is unduly distressed by meeting dependency needs; or the mother who sees her child as always lying has been made to feel guilty about minor transgressions as a youngster. Another line of research is to follow the consequences of the mother's distortions in the child's development. At a simple level there may be a direct relation between distortion and development; the healthy infant who is seen as "always sick" by his mother eventually might become sickly, or an infant who is seen as "mean from birth" may, in fact, become mean. At a more complex level, the child might develop the exact opposite characteristics the parent professes since he responds to strong covert or unconscious needs. In light of the richness of clinical evidence, there is a surprising lack of experimental investigation of distorted perceptions, their relations to other aspects of the mother's personality and history, and their effects on the child's development. In such research, the nondistorting mothers can serve as a ready-made control group.

In regard to "valid" mothers, there is evidence (3) which suggests that they are not necessarily "good" mothers. A mother can correlate highly with experts in evaluating early development, she can also evidence a good intellectual awareness of her child's current thinking, and yet behave in a damaging manner, partly because of a lack of positive feeling and partly because she uses her understanding to manipulate the child so as to achieve her own ends. One can speculate that, eventually, there will be evidence of "invalid" mothers who are genuinely maternal. These would be women who function at an intuitive, affective level and who have little interest in the kind of conceptualization and reflection necessary to evaluate themselves and their children. Certainly, we need to know more about the relation between "awareness" and maternal behavior since the association does not appear to be a simple one.

Fidelity of Memory

Just as in the previous section it was important to understand the objective situation and the mother-as-perceiver, it is important here to understand the subsequent history of the child and the nature of the mother-as-rememberer.

Subsequent events might be conceptualized as interpolated material and, as is well known, it is the nature of this material rather than the passage of time per se which crucially affects recall. However, it is difficult to hypothesize what the precise nature of its influence will be. Take an hypothetical example of two 4-year-olds with a similar history who attend preschool for a year. One adjusts well, makes friends, discovers new skills, and generally sets up positive reverberations within the family; the other one has separation anxieties, becomes a behavior problem, and increases tension within the family. What will be the differential effect upon recall of early development? Will one mother be set to remember good things, and the other bad? Will continuity between the immediate and remote past decrease accuracy of recall because of the relative homogeneity of events (i.e., everything going well or everything going badly), while discontinuity heightens accuracy through contrast?

Such questions are difficult enough assuming a kind of normal memory in all mothers. However, the mother's particular intellectual and personality characteristics introduce another set of variables. One would expect a compulsive individual who keeps a baby book and compares her child's development with Gesell's norms to withstand the inroads of all kinds of subsequent events better than an hysterical one who lives from day to day and relies on repression to protect her from painful memories. Paul's experiments (7), which are conceptualized in terms of "cognitive style," also furnish clues regarding different kinds of recall. Using a Bartlett-type situation, he distinguished individuals who import new material from ones who skeletonize material when they recall and further differentiated importations which make the material more coherent from those which merely embellish it. Although such findings are somewhat limited at present, the approach promises to be a valuable supplement to our knowledge of the effect of defense mechanisms and personality disorders upon memory.

The case in which the mother's subsequent experiences increase the validity of her evaluation of early events is of special interest. Change does not have to be in the direction of repression, distortion, and defensiveness, but it can be in the direction of greater clarity of understanding. Having more children, for example, can make a mother realize that her first born was not as fretful and demanding as she had thought. One might speculate that this capacity to reevaluate the past more realistically is one measure of the mother's capacity for psychological growth.

Only a series of empirical studies can determine which of the above variables is significant and in what way they affect memory. Some of these studies can be incorporated into the design of longitudinal projects with relative ease. Macfarlane, in the previously cited investigation (6), has made a start in this direction. Her finding that mothers tend to err in the direction of precocity suggests that, in general, distance lends enchantment to the neonatal period. However, her reference to the complex nature of the data hints that there may be a number of specific relations which must be teased out.

It also may be that the time span studied can be a relatively short one rather than the entire period of childhood. If forgetting here follows the same course that it does in other memory research, about as much information might be lost after two years as after ten. The author is presently conducting a study which compares the histories of the first three years of life given by mothers of 5-year-old and 9-year-old children. Factual information comes from detailed medical records which have been kept since birth on the children. Taking a cue from Macfarlane and the experimental literature, it is hypothesized that there will be only a slight increase in unreliability in the 9-year-old group as compared with the 5-year-old. The data will be further analyzed to ascertain what kinds of information are reliably and unreliably remembered and what kinds of mothers are particularly accurate and inaccurate in their recall.

RECALL UPON REQUEST

The crowning achievement of the Ideal Research Mother is that the total past is totally available upon request. With more fallible women, recall and reporting of information is, to a certain extent, a function of the method used in trying to obtain it.

In comparison with the extensive literature on the art of interviewing parents, the number of systematic studies is relatively small. Fortunately, the technique is shared by a number of disciplines, and there is a substantial body of research emanating especially from social psychology. Many controversies are not settled, but important variables have been isolated and studied, and the psychologist has at his disposal valuable orienting ideas as to the consequences of choosing one technique over the other. For instance, the advantages and disadvantages of the structured and unstructured interview have been spelled out in sufficient detail to allow the interviewer to judge which is more appropriate for his purposes and conduct systematic studies of the two procedures in eliciting historical material. Antonovsky (2) used structured interviews, unstructured interviews, and observation to test a number of hypotheses about childrearing practices and the behavior of children. Of the three techniques, the structured interview was the most satisfactory in that it yielded more correlations in line with theoretical predictions. Further analysis revealed that, contrary to prediction, there was no closer relation between the two sets of interview data than between interviews and observations. In order to account for this, Antonovsky hypothesized that the wide variability on the three techniques which was noted in certain mothers was due to greater anxiety and conflict in them. Again, this hypothesis was not supported. Although such a study raises more questions than it answers, it represents a rare attempt to evaluate a technique and to discover the interaction between techniques and personality variables.

There is also increasing interest in and information about the interpersonal aspects of obtaining information. The mother can no longer be regarded as an eager, passive recipient, waiting to answer all questions to the best of her ability. Instead, she has her own concern about the picture of herself and her child which she presents to the interviewer, as well as her own reactions to his personality. There is ample evidence that the pull toward social acceptability is great, and this problem plagues interviewers and inventory makers alike. The mother's view of the purpose of the interview may be another selective factor, and a developmental history of a child with an ulcer may differ depending on whether it is given to a physician or a social worker.

Finally, the individual characteristics the interviewer brings to his role probably influences the response of the mother. Again drawing from research in other fields, there is evidence that his class, race, sex, and status significantly affect the kinds of attitudes expressed. Furthermore, respondents tend to say they share their interrogator's views on politics, race relations, war, and prefabricated houses. A similar phenomenon may exist in child psychology, where the interviewer's anxieties about specific areas might subtly distort his questioning and recording or cue a defensive reaction on the part of the mother. Fortunately, there is also evidence that such biases in the data can be controlled by training.

To answer the questions raised in this brief outline would take a major research effort. One might even wryly conclude that, by the time we have sufficient information to validate the mothers' histories, we will know enough about child development to eliminate her as a source of information. This may be so, and the reconstructive approach may eventually wither away or become a technique of convenience rather than a source of basic data. In the meantime, there is no reason to proceed quite as blindly as we have been. The individual studies suggested above will give a clearer idea of what kinds of events can be validly observed and remembered or what kinds of mothers give valid and reliable histories.

The Factual vs. the Subjective Approach

The discussion so far has centered on the question, Is what the mother reports really true? It can be argued that the question itself is irrelevant. Certain personality theories and schools of psychotherapy maintain that it is not reality which matters but the individual's perception of reality. Thus, when a mother says, "The baby was always crying," what is important is the mother's evaluation of the infant as a constant crier and the attitudes of exasperation or desperation which accompany such a perception. It could further be argued that it is just such perceptions, with their distinctive cognitive and affective components, which are the crucial formative influences in the child's life and most worthy of study.

This is no place to discuss the merits of the subjective approach, but the methodological problems it involves should be touched upon briefly.

In one sense, such problems are no different from those of the factual approach. In regard to reliability, the question, Did this really happen? becomes, Did the mother really perceive the event the way she said she did? It is reasonable to assume that evaluations of situations change with time and that an experience takes on a different affective coloring. A pilot study being conducted by the author indicates that recall of feeling is much less reliable than recall of fact; e.g., a mother usually does not change her story about whether she breast or bottle fed her infant, but the passage of time significantly affects her report of how she felt about this activity. If substantiated, this finding would raise serious doubts about relying exclusively on subjective material to reconstruct the significant influences in the child's development.

In regard to validity, it seems likely that the subjectivist-researcher will not be able to dismiss reality so readily as his clinical counterpart. One of the most baffling questions in studying the etiology of childhood psychosis, for example, is whether the mother's deviant attitudes cause the pathology or represent reactions to a constitutionally deviant development. One wonders what kinds of lawfulness would be discovered if mothers were grouped together only on the basis of the similarity of their subjective reactions; it would be surprising if the realistically and the neurotically determined perceptions had the same effects on the child's development.

There are two ways for the subjectivist to avoid the methodological problems under discussion. One is to take an ahistorical tack and claim that it is only the mother's present report, only her present perception of her perceptions, which is important. The past is relevant only as it influences the present. This is a defensible position and one suitable to the study of current parent-child relations. However, it automatically eliminates the mother's history as data on which to reconstruct her child's early life.

The other solution is to leave the reconstruction in the hands of experts. These specially qualified individuals could decide whether the mother is an accurate and truthful person, her areas of distortion, how well her report reveals her true feelings, and how much it is a defense against underlying attitudes. This may settle one set of problems—namely, those concerned with, What are the facts?—but it raises a host of others concerning the nature and reliability of expert judgment. At this point, it becomes a matter of opinion as to whether the cure is worse than the disease.

Fruitfulness and Criterion Validity

Perhaps, one reason why there have been so few criterion validity studies of the interview is that it has been so fruitful. The supreme example is Freud's use of such material to evolve theories and hypotheses which have supplied a generation of developmental psychologists with research ideas. Or, if one would object to regarding a psychoanalysis as merely a series of interviews, the more modest work of Levy (5) on maternal overprotection or Sears (8) on patterns of childrearing can be cited. As long as interview material reveals consistent relationships within itself (e.g., severe toilet training by cold mothers increases the amount of upset in the child), as long as reported antecedents are consistently found for certain behaviors (e.g., overprotected children have a history of excessive contact with the mother during infancy), and as long as such consistencies either test hypotheses or lead to further hypotheses, the interview justifies its extensive use.

There certainly can be no quarrel with such a defense. Yet it is also true that, without at least some of the methodological studies suggested above, one can never be sure of the meaning of either positive or negative findings. For instance, Gerard and Siegel (4), using interviews of schizophrenic and normal controls, found a number of significant differences but, what is more to the point, also a number of nonsignificant differences in areas where they would be expected, such as punitiveness of the parent, feeding disturbances, and frequency of behavior problems. At this point, the investigator is hopeless to know whether negative results were obtained because in fact no differences exist between the two groups or because the interview was an insensitive or an inappropriate technique for revealing such differences. Another case in point: Altman (1) found no correlation between judges' ratings of maternal attitudes based on interview material and certain personality characteristics of normal children. However, when the interviewer rated the mothers on the same characteristics, positive relations were found. Thus, there is a suggestion that the interviewer makes the most sensitive judge, since the traditional method of evaluating the data obscured rather than revealed important differences. The more general point is that negative findings will continue to have an ephemeral and arbitrary quality until we are more certain of the nature of our measuring instrument.

The same point holds for positive findings. At present, there is no way of being sure whether the obtained relations are between actual historical events, between perceptions of events, or between perceptions of events and reality. For instance, a mother's memory of her feelings about breast feeding may be completely distorted and yet correlate highly with behavior deviations in her child if a common personality dynamic is responsible for both. At the methodological level, the interviewer who knows the kinds of questions which are likely to elicit unreliable answers and the types of mothers who are likely to distort developmental data will not have his faith in his technique shaken by evidence that reported events did not really happen. The more we can distinguish those parts of a history which represent accurate reporting from those which represent distortions, the more certain we can be about the nature of the relations which are found and the more satisfying the interview will become as a research instrument.

REFERENCES

1. ALTMAN, C. H. Relationships between maternal attitudes and child personality structure. *Amer. J. Orthopsychiat.*, 1958, 28, 160-169.
2. ANTONOVSKY, H. F. A contribution to research in the area of the mother-child relationship. *Child Develpm.*, 1959, 30, 37-51.
3. GARNER, A. M., & WENAR, C. *The mother-child interaction in psychosomatic disorders.* Univer. of Illinois Press, 1959.
4. GERARD, D. L., & SIEGEL, J. The family background of schizophrenia. *Psychiat. Quart.*, 1950, 24, 47-73.
5. LEVY, D. M. *Maternal overprotection.* Columbia Univer. Press, 1943.

6. Macfarlane, J. W. Studies in child guidance. I. Methodology of data collection and organization. *Monogr. Soc. Res. Child Develpm.*, 1938, 3, No. 6.
7. Paul, I. H. Studies in remembering. *Psychol. Issues,* 1959, 1, No. 2.
8. Sears, R. R., Maccoby, E. E., & Levin, H. *Patterns of child rearing.* Row, Peterson, 1957.
9. Smith, H. T. A comparison of interview and observation measures of mother behavior. *J. abnorm. soc. Psychol.,* 1958, 57, 278-282.

FAMILY RELATIONS OF BRIGHT HIGH-ACHIEVING AND UNDER-ACHIEVING HIGH SCHOOL BOYS[1]

William R. Morrow*
Fulton (Missouri) State Hospital

and Robert C. Wilson
Portland (Oregon) State College

The present article reports data on the family relations of bright high school boys making good grades as compared with bright high school boys making mediocre or poor grades.

Of numerous studies of school achievement, few have obtained data regarding family relations correlates. These few have indicated a positive association between student achievement and emotionally supportive home situations. This general finding applies to elementary school pupils (4, 9, 15, 18, 20,), high school students (13, 14), and college students (12).

More specifically, parents of high-achievers have been found to give their children more praise and approval (18), to show more interest and understanding (19), to be closer to their children (14), to make their children feel more family "belongingness" (20) and identification with parents (19). On the other hand, parents of under-achievers have been reported to be more domineering (12, 14) and overrestrictive (18) and to use more severe and frequent punishment (4, 14), which is at the same time less effectual (4). Parents of under-achievers have also been found more likely either to baby their youngsters *or* to push them excessively (9) and to present to their

---

* Missouri State Hospital No. 1, Fulton.

[1] The research reported here was part of a larger study of bright high- and under-achieving high school boys, sponsored by the Gifted Child Project of the Portland, Oregon, Public Schools and supported by a grant from the Fund for the Advancement of Education.

Child Develpm., 1961, 32, 501-510.

youngsters either low *or* extremely high (pressuring) demands for achievement (18). Finally, homes of under-achievers are reported to show more tension (12) and more parental disagreement as to standards of behavior expected of their youngsters (19).

Most of these studies have been based on qualitative data such as interviews. Many had sampling limitations such as a very small sample and questionable equation of groups. No previous study seems to have dealt with family relations correlates of school achievement among high school students of superior intelligence.

In the present research the groups compared were carefully equated, the sample was relatively homogeneous in intelligence, all subjects being of superior intelligence, and the data were obtained by structured measuring instruments.

Before the present study was initiated, exploratory interviews covering various topics including family relations were held with a few bright high- and under-achieving high school boys.[2] On the basis of these interviews as well as the studies cited, it was hypothesized that the reported family relations of high-achievers, as contrasted with those of under-achievers, are characterized by: (a) more emotionally supportive home environments; and, more specifically, (b) greater family sharing in recreation, decision-making, and exchange of confidences and ideas; (c) greater mutual parent-child affection, acceptance, trust, and approval; (d) less parental domination, severity, and restrictiveness; (e) more sympathetic parental encouragement of achievement, but less overinsistence on achievement; (f) greater harmony between parents and more regularity of home routines.

It was hypothesized further that students' family relations influence their school achievement via certain mediating variables. Specifically, it was hypothesized that supportive home environments are associated with positive student attitudes toward (a) teachers, (b) school, and (c) intellectual activities, as representing the adult world of parents, and that each of these mediating attitude variables would be associated with school achievement.

METHOD

Sample

The sample included two equated groups each containing 48 high school boys of superior intelligence (120 IQ or above). The groups were equated for grade in school, socioeconomic status,[3] and intelligence. Each group contained 19 ninth-graders, 14 tenth-graders, and 15 eleventh-graders. Each group likewise contained one upper class, 15 upper-middle class, 23 lower-middle class, and 9 lower class boys. The mean IQ of the high-achieving group was 126.0; that of the under-achievers, 125.3, the difference being nonsignificant.

---

[2] These interviews were conducted by Morrow, using a schedule which he prepared.

[3] After all other criteria had been met except equating the groups for socioeconomic status and IQ, the high-achieving group included five more upper class boys than did the under-achieving group, 12 more upper-middle class boys, and nine more lower-middle class boys. The under-achievers included 12 more lower class boys than did the high-achieving group. This imbalance is consistent with previous findings of a positive association between academic achievement and family socioeconomic status for elementary and high school students (2, 3, 5, 16, 17), though not for college students, for whom findings on this point have been inconsistent [*see*, for example, (8)].

The groups differed in grade-point average in academic courses (with a minimum of three academic courses for each student). The high-achievers maintained an average of 1.00 to 1.67 (1 being the best grade possible, 5 the poorest) during the school year 1955-1956. The under-achievers maintained an average of 2.75 to 5.00, 2.75 being regarded as under-achievement for students of superior intelligence.

Data Collection

Several types of information about family relations were obtained from the students through group-administered questionnaires. The students were not asked to sign their names, but birthdates and other data were used to identify questionnaires for the purpose of selecting the sample. Otherwise, anonymity was maintained and information on individuals kept confidential.

The students' family relations as seen by themselves were evaluated primarily by 16 self-report Family Relations Scales. Each scale consisted of six questions about the student's relations with his parents (or foster parents). The scales were presented in consecutive order, but without scale titles or breaks in spacing. Each student was asked to indicate to what extent each item described his own home situation, using the following four response categories:

1. Not at all or Almost Never
2. A Little or Sometimes
3. Considerably or Often
4. Very Much or Very Often (or Almost Always)

The student was assigned a score on each scale by summing his scores on the six items in the scale. A score was also obtained on the total of all scales (except Harmony of Parents[4]), conceived as an index of Over-all Family Morale.

The general form of the scales, as well as much of their content, was adapted from a questionnaire developed by Brown and his associates (10, Ch. 21).[5] Seven scales were added; and a number of Brown's items were omitted, revised, or supplemented by additional items. Other sources of item content included instruments developed by Block (1), Itkin (11), Rickard (18), Stott,[6] and Strodtbeck,[7] as well as our preliminary interviews.

---

[4] The Harmony of Parents Scale was scored only for students with two parents living together. The *N* for this scale was 40 in each achievement group, re-equated for grade in school, socioeconomic status, and intelligence.

[5] We wish to thank Dr. Robert J. Havighurst for making available to us a copy of the Family Relations Questionnaire used in his research. That questionnaire was developed by Dr. Andrew Brown and his associates for use in a larger study of adolescent character and development, directed by Dr. Havighurst and Dr. Hilda Taba.

Following are the scale titles and a sample item for each scale:

1. *Family Sharing of Recreation.* "Do you and your family go on picnics or outings or trips together?"
2. *Family Sharing of Confidences and Ideas.* "Do your parents discuss their work and activities with you?"
3. *Family Sharing in Making Decisions.* Do your parents let you help decide everyday family policies, rules, and ways of living?"
4. *Parental Approval.* "Does either parent ever seem to wish that you were a different sort of person?" (Negative item, scored in reverse.)
5. *Parental Affection.* "Do your parents openly show affection for you by word or action?"
6. *Parental Trust.* "How confident do your parents seem to be that you will behave properly away from home?"
7. *Parental Approval of Peer Activities.* "Do your parents object to some of your activities with your friends and acquaintances?" (Negative item.)
8. *Student Acceptance of Parental Standards.* "Do you agree with your parents' ideas about life?"
9. *Student Affection and Respect for Parents.* "Would you like to be the same kind of parent that your parents have been?"
10. *Lack of Parental Overrestrictiveness.* "Do your parents try to direct your activities?" (Negative item.)
11. *Lack of Parental Severity of Discipline.* "How often do your parents punish you?" (Negative item.)
12. *Lack of Parental Overprotection.* "Do your parents try to protect you too much against difficulties or dangers?" (Negative item.)
13. *Lack of Parental Overinsistence on Achievement.* "Are your parents always after you to work hard to become a success?" (Negative item.)
14. *Sympathetic Encouragement of Achievement.* "Do your parents inspire you to want to develop your abilities?"
15. *Regularity of Home Routine.* "Are your meals served at regular hours?"
16. *Harmony of Parents.* "Do your parents openly show affection or consideration for each other?"

The students were also asked to provide sociological data on parents' marital status, occupation, and education and on the ages and sexes of their siblings. In addition, each student was asked to respond to three checklists: (a) a list of activities (e.g., "music," "insects or beach life," "fish or hunt," "billiards") to be checked as to which ones his parents had taught him or developed his interest in, yielding an "intellectual," a "nonintellectual," and a total score; (b) a list of 11 high school goals or values (in the areas of academic achievement, athletics, social relations and status, general adjustment, morals, and religion) to be checked as to which his parents considered "most" and "least" important (three each to be checked) for him to attain; (c) a similar list of 11 adult goals (e.g., "be outstanding in your occupation," "live a happy life," etc.) to be checked the same way.

---

[6] Unpublished Home Life Scale, developed for use in a study of family morale conducted by Dr. Leland H. Stott. We wish to thank Dr. Stott for making this instrument available to us.

[7] Unpublished Youth Planning Survey questionnaire, developed by Dr. Fred L. Strodtbeck and his associates for use in a study of achievement motivation and its correlates. We wish to thank Dr. Strodtbeck for making a copy of this questionnaire available to us.

The questionnaire also included four open-ended questions: "What sort of person is your mother?" "What sort of person is your father?" "What do you like most about your home and family?" "What would you like to change about your home and family?"

The following measures of hypothesized mediating variables (between family relations and school achievement) were obtained: (a) a six-item attitude scale (with a five-point scale for each item) designed to measure Negative Attitudes to Teachers,[8] which yielded a corrected odd-even reliability of .69; (b) a similar six-item attitude scale designed to measure Negative Attitudes to School,[8] which yielded a corrected odd-even reliability of .85; (c) a 72-item interest-inventory scale (with dichotomous items) designed to measure Interest in Intellectual Activities[9] (esthetic activities, social problems, natural science, and formal symbol manipulation such as mathematics and puzzles), which yielded a corrected odd-even reliability of .77.

Analysis of Data

The internal consistency reliability of each Family Relations Scale and of the Over-all Family Morale Scale was evaluated by computing an odd-even product-moment reliability coefficient. Differences between high- and under-achievers on each scale and on the Over-all scale were tested for significance by a median test.

Group differences in checklist responses were tested for significance by a median test for scale scores and by a chi square test for single items.

As for the open-ended questions,[10] response categories potentially differentiating between the two achievement groups were determined by studying an inspection sample. These categories were cross-validated by being scored[11] "blind" on another sample of 66[12] boys in each group with groups equated for grade in school, socioeconomic status, and intelligence. Differences between groups were tested for significance by chi square and median tests.

Pearson product-moment correlations were obtained between the Over-all Family Morale scale and the several measures of mediating variables concerned with attitudes toward teachers, school, and intellectual activities. These correlations were computed separately for the total sample, for high-achievers only, and for under-achievers only. In addition, differences between achievement groups on these mediating variables were tested for significance by a median test.

---

[8] Adapted from the Brown-Holtzman Survey of Study Habits and Attitudes (S.S.H.A.), published by the Psychological Corporation. We are grateful to the Psychological Corporation for permission to make use of the S.S.H.A. for this purpose.

[9] Adapted from an inventory developed by J. P. Guilford and his associates (6, 7). We are grateful to Dr. Guilford for making the inventory available to us for this purpose.

[10] The analysis of the open-ended questions in the student questionnaire was conducted by J. Robert Wallace under Morrow's supervision.

[11] In view of the low validity obtained in this cross-validation, time was not invested in having a second judge score the material to provide a measure of interjudge agreement.

[12] This sample included 30 additional 1956-1957 freshmen in each group.

All tests of significance used were two-tailed.

RESULTS

Family Relations Scales

Results on the Family Relations Scales are presented in Table 1. The internal consistency reliability of the 16 six-item scales appeared to be sufficiently satisfactory for purposes of group comparison: all but three of the corrected odd-even coefficients were above .70. The 90-item Over-all Family Morale Scale yielded a reliability of .97.

TABLE 1

FAMILY RELATIONS SCALES: MEDIAN TESTS*

		PER CENT ABOVE MEDIAN		
Scale Title	r†	*Highs* ($N=48$)	*Lows* ($N=48$)	p
Family Sharing of Recreation	.76	69	44	.02
Family Sharing of Confidence & Ideas	.84	63	35	.01
Family Sharing in Making Decisions	.88	60	44	*ns*
Parental Approval	.56	73	33	.001
Parental Affection	.88	60	42	*ns*
Parental Trust	.73	60	25	.001
Parental Approval of Peer Activities	.94	71	42	.01
Student Acceptance of Parental Standards	.69	52	25	.01
Student Affection & Respect toward Parents	.91	58	44	*ns*
Lack of Parental Overrestrictiveness	.63	56	29	.01
Lack of Parental Severity of Discipline	.70	69	42	.01
Lack of Parental Overprotection	.77	52	56	*ns*
Lack of Parental Overinsistence on Achievement	.75	63	46	*ns*
Parental Encouragement of Achievement	.74	60	40	.05
Harmony of Parents ($N=40$)	.72	63	48	*ns*
Regularity of Home Routine	.76	52	46	*ns*
Over-all Family Morale	.97	67	33	.001

* Two-tailed tests.

† Odd-even reliability coefficient, corrected by Spearman-Brown formula.

Nine of the 16 six-item scales differentiated between the high- and under-achieving groups in the predicted direction beyond the .05 level, seven of these nine beyond the .01 level. The Over-all Family Morale Scale differentiated beyond the .001 level.

High-achievers more often than under-achievers (a) described their families as typically sharing recreation, ideas, and confidences; (b) described their parents as approving and trusting (the areas of sharpest difference between the two groups), affectionate, encouraging (but not pressuring) with respect to achievement, and relatively nonrestrictive and nonsevere; and (c) described themselves as accepting their parents' standards.

An equal majority of both groups described their parents as having a relatively harmonious relationship, portrayed their homes as having a fairly regular routine, denied that they were either seriously overprotected or excessively pressured to achieve, and said they felt considerable respect and affection for their parents.

Even in the areas of greatest difference, however, there was considerable overlap between the two groups. On the Over-all Family Morale Scale nearly a third of the high-achievers scored below the median, and nearly a third of the under-achievers scored above the median.

Sociological Data

The two groups did not differ significantly in any of the sociological factors on which data was obtained. The data contradict the stereotyped notion that a mother's working outside the home inevitably leads to neglect which conduces to poor school performance (and other dire consequences). Actually 47 per cent of the high-achievers (as against 37 per cent of the under-achievers) reported that their mothers were working outside the home!

Check Lists

The two groups showed essentially no significant differences in their responses to the check lists.

Open-Ended Questions

In the inspection sample 34 potentially differentiating response categories were noted, of which 25 were grouped in five "high" combined categories and nine were grouped in three "low" combined categories. In the cross-validation only two "high" combined categories (out of eight combined categories) and four single "high" categories differentiated significantly (in the predicted direction).

Fifty-one per cent of the high-achievers as against 33 per cent of the under-achievers ($p < .05$) were scored in the combined category, "Positive references to intrafamily relationships." Two single categories in this grouping also differentiated significantly: "References to parental interest in family or student" (29 per cent to 12 per cent, $p < .01$); and "References to parents' outgoing, positive shaping of student's development" (12 per cent to 0 per cent, $p < .02$).

The combined category, "References to 'Golden Rule' virtues of parents" was scored significantly more often for high-achievers (56 per cent to 29 per cent, $p < .01$). Two single categories in this grouping also differentiated significantly: "Parents described as 'considerate' or 'thoughtful'" (17 per cent to 4 per cent, $p < .05$); and "Parents described as 'understanding'" (36 per cent to 17 per cent, $p < .02$).

These results are consistent with those on the Family Relations Scales.

Mediating Variables

The correlation between Over-all Family Morale and Negative Attitudes to Teachers was −.67 for the total sample, −.46 for high-achievers only, and −.60 for under-achievers only. The corresponding correlations between Family Morale and Negative Attitudes to School were −.69, −.32, and −.65. The correlations between Family Morale and Intellectual Interests were .57, .31, and .49. All of the obtained correlations are significant at the .05 level or better. On each of the three measures of mediating variables the two achievement groups differed at the .001 level in the expected direction.

These results are consistent with the hypothesis that supportive family relations foster academic achievement *via* promoting positive attitudes toward teachers, school, and intellectual activities, as symbols of the adult world of parents. However, other interpretations of the direction of causality are not excluded.

SUMMARY

The reported family relations of 48 high school boys of superior intelligence making high grades were compared with those of a group making mediocre or poor grades, equated for grade in school, socioeconomic status, and intelligence. The main measuring instrument was a set of 16 six-item questionnaire scales (with a four-point scale for each item), on which each subject was asked to describe his family relations. The students also provided sociological data, check-list data on parental goals for the student, and open-ended question data on conceptions of parents. In addition, attitude scale data were obtained on variables hypothesized to mediate the influence of family morale on student achievement.

The results supported the hypotheses that bright high-achievers' parents reportedly engage in more sharing of activities, ideas, and confidences; are more approving and trusting, affectionate, and encouraging (but not pressuring) with respect to achievement; are less restrictive and severe; and enjoy more acceptance of parental standards by their youngsters. Not supported were hypotheses that under-achievers' families show more overprotectiveness, more high-pressure for achievement, more parental disharmony, more irregularity of home routine; differences in goals for their youngsters; or differences in sociological factors such as parents' marital status, current occupation of either parent, or number and ages of siblings.

The results also supported the hypothesis that family morale fosters academic achievement among bright high school boys *via* fostering positive attitudes toward teachers and toward school and interest in intellectual activities, as mediating variables. However, other hypotheses as to the direction of causality are not ruled out.

REFERENCES

1. BLOCK, V. L. Conflicts of adolescents with their mothers. *J. abnorm. soc. Psychol.*, 1937, 32, 193-206.
2. COLEMAN, H. A. The relationship of socio-economic status to the performance of junior high school students. *J. exp. Educ.*, 1940, 9, 61-63.
3. COLLINS, J. H., & DOUGLAS, H. R. The socio-economic status of the home as a factor in success in the junior high school. *Elem. Sch. J.*, 1937, 38, 107-113.
4. CONKLIN, A. M. Failures of highly intelligent pupils. *Teach. Coll. Contrib. Educ.*, 1940, No. 792.
5. GOUGH, H. G. The relationship of socio-economic status to personality inventory and achievement test scores. *J. educ. Psychol.*, 1946, 37, 527-540.
6. GUILFORD, J. P., CHRISTENSEN, P. R., BOND, N. A., JR., & SUTTON, M. A. *Technical appendices to the research bulletin, A factor-analytic study of human interests.* Beverly Hills, Calif.: Sheridan Supply Co., 1953.
7. GUILFORD, J. P., CHRISTENSEN, P. R., BOND, N. A., JR., & SUTTON, M. A. A factor analysis study of human interests. *Psychol. Monogr.*, 1954, 68, No. 4.
8. HARRIS, D. Factors affecting college grades: a review of the literature, 1930-1937. *Psychol. Bull.*, 1940, 37, 125-161.
9. HATTWICK, B. W., & STOWELL, M. The relation of parental over-attentiveness to children's work habits and social adjustments in kindergarten and the first six grades of school. *J. educ. Res.*, 1936, 30, 169-176.
10. HAVIGHURST, R. J., & TABA, H. *Adolescent character and development.* Wiley, 1949.
11. ITKIN, W. Some relationships between intra-family attitudes and preparental attitudes toward children. *J. genet. Psychol.*, 1952, 80, 221-252.
12. JONES, E. S. The probation student: what he is like and what can be done about it. *J. educ. Res.*, 1955, 49, 93-102.
13. KIMBALL, B. The sentence-completion technique in a study of scholastic achievement. *J. consult. Psychol.*, 1952, 16, 353-358.
14. KIMBALL, B. Case studies in educational failure during adolescence. *Amer. J. Orthopsychiat.*, 1953, 23, 406-415.

15. LEIBMAN, O. B. The relationship of personal and social adjustment to academic achievement in the elementary school. Unpublished doctoral dissertation, Columbia Univer., 1953.

16. LEWIS, W. D. A comparative study of the personalities, interests, and home backgrounds of gifted children of superior and inferior educational achievement. *J. genet. Psychol.*, 1941, 59, 207-218.

17. MUSSELMAN, J. W. Factors associated with the achievement of high school pupils of superior intelligence. *J. exp. Educ.*, 1942, 11, 53-68.

18. RICKARD, G. The relationship between parental behavior and children's achievement behavior. Unpublished doctoral dissertation, Harvard Univer., 1954.

19. TIBBETS, J. R. The role of parent-child relationships in the achievement of high school pupils. *Dissert. Abstr.*, 1955, 15, 232.

20. WALSH, A. *Self-concepts of bright boys with learning difficulties.* Columbia Univer., 1956.

PEER POPULARITY AND DEPENDENCE ON ADULTS IN PRESCHOOL-AGE SOCIALIZATION

Boyd R. McCandless,*
State University of Iowa

Carolyn B. Bilous *and* Hannah Lou Bennett
University of Hawaii

The present research was designed to investigate the interaction between popularity, dependency on adults, and sex in preschool-age children. Marshall and McCandless, studying 38 preschool children in two groups of the Iowa Child Welfare Research Station's Laboratory Preschools, found that "dependence on adults in the preschool situation accompanies relatively low social status and participation" (2, p. 419). In that investigation dependency was defined as simply the total number of interactions with adults during a prescribed period of observation. In addition to this over-all negative relation between popularity and dependence, Marshall and McCandless found that sex differences significantly characterized the interaction between popularity and dependence; girls showed a median correlation between popularity and dependence of about —.70, while the median correlation for boys was about —.40.

The present study, then, began as an investigation of why (and if) dependency interferes more with popularity for girls than boys.

The correlations reported by Marshall and McCandless imply that a relatively high percentage of adult contacts involve some sort of socially inap-

---

*Iowa Child Welfare Research Station, State University of Iowa, Iowa City; on leave, University of the Panjab, Lahore, West Pakistan.

Child Develpm., 1961, 32, 511-518.

propriate dependency and that the girl-adult contacts are more characterized by this than the boy-adult contacts. From observation of preschool play it seemed likely that one frequent type of socially inappropriate dependency contact with adults is seeking adult intervention to solve peer-conflict situations. It was hypothesized that seeking adult intervention in peer-conflict situations would interfere with popularity and that a higher percentage of the adult contacts would be comprised of attempts to seek adult intervention in peer-conflicts for girls than for boys. This would in part account for the differential tendency for dependency to interfere more with popularity for girls than for boys.

Heathers (1) has suggested that the total number of contacts with adults can be usefully classified into two separate categories: (a) emotional dependency and (b) instrumental dependency. Such a classification seemed useful for the present investigation. It is here suggested that children who are emotionally dependent upon adults are motivated to learn social skills necessary to gain emotional satisfactions from adults, rather than to learn skills necessary to gain emotional satisfactions from peers. Thus, a differential tendency for girls more than boys to be *emotionally* dependent upon adults may also account for the previous findings that dependency interferes more with popularity for girls than for boys. Instrumental dependency, on the other hand, since it is unrelated to interpersonal, emotional need-satisfaction, should not interfere with popularity.

It would also be expected that the differential responses of girls and boys to mother-figures (teachers) had been learned in the home. Thus, it was expected that the mothers of girls would more frequently act as problem solvers in conflict situations and be more nurturant and supportive of the child in conflict situations than would the mothers of boys.

OBTAINED MEASURES

Popularity

Teacher judgments. The head teacher and her three student assistant teachers of each of the three preschool groups involved in this study were asked to nominate the four best friends of each child in the respective groups. Thus, a total of 16 nominations for each child were used. The index of popularity used was the simple count of number of mentions as "one of the four best friends" of other children. This was then converted to a ratio of possible mentions. Average interteacher correlation was .54.

Free play observations. As an alternative (and probably sounder) index of popularity, the time spent during observation by each child in associative and cooperative play with every other child was tabulated. The best friends of each child were considered to be the four with whom he spent the most time in such contact. That child who was one of the four best friends of the most children was designated as most popular; the one who least frequently was one of the four best friends of others was designated least popular. Thus, popularity was defined as "being sought after" by peers.

Child Behavior with Adults (Teachers)

Fifteen five-minute samples of data for each child were obtained by the first author according to a semiformal time sample observational technique developed by him during observations conducted prior to the beginning of the present study. Observational focus was on peer conflicts and their solution, although notations of a number of social interactions of the child during each five-minute period were also made. Reliability of observations between the first and third authors for approximately four hours of observation was 92 per cent exact agreement of categorization. As a further check on the technique, percentage of agreement between the first author and a psychology graduate student who had no prior experience with observation of children was 65 during an initial two-hour observation and, after discussion, rose to 93 for a second and subsequent two-hour observational period.[1]

For the present study, the following observed events were tabulated for each child: (a) number of contacts made with adults, (b) total time spent interacting with adults, (c) number of times instrumental "nurturance" was solicited from adults, (d) number of times emotional "nurturance" was solicited from adults, and (e) the number of times that a child requested or threatened to request help from the teacher.

Mother Behavior with Child

Since it was not feasible actually to obtain observational data concerning mother-child interactions, reliance was placed upon situation-bound maternal interviews. Mothers were asked to describe the last two instances of conflict in which they had observed their children and to tell what they were about, who was involved, what had gone on, and what their role had been. If possible, one of these conflicts was between the child and his siblings, while the other conflict involved a neighborhood playmate or playmates. For only children, or children who had only substantially older siblings, two neighborhood conflicts were elicited. The second and third authors conducted all interviews and secured excellent cooperation from the mothers. From typescripts of the interviews 29 categories of maternal response to child conflict were tabulated. The three authors classified 10 interviews with full discussion of their disagreements, then independently coded the remaining interviews. Percentages of exact agreement on categorization were: between authors 1 and 2, 81.4; 1 and 3, 70.6; and 2 and 3, 75.5.

Two clinical type scales were then devised to lump the categories for purposes of the present investigation: (a) *"Let Them Fight Their Own*

---

[1] The authors appreciate the work of Mr. Mark Beals in establishing reliability of observations.

Battles." This scale lumps together those maternal responses where there is no active maternal intervention. Sample behaviors are: encourages child to defend self, ignores the issue or situation, encourages child to work through own conflicts. (b) "*Active Intervention with Constructive Intent.*" Sample items are: suggests compromise with explanation and/or reasoning, suggests withdrawal from conflict for ethical reasons, encourages child to come to mother.

SUBJECTS

The subjects (*Ss*) were 26 Caucasian youngsters, referred to in Hawaii as ***haoles***, and 36 ***nonhaole*** children. The latter were of predominantly Oriental ethnic origin. Two children—one a blind girl, the second a mentally retarded boy—were dropped from the study, leaving a final *N* of 60, 28 boys and 32 girls, of whom 26 were ***haole***. Ages ranged from 3 years, 6 months to 5 years, 3 months, the mean being 4 years, 5 months. All children attended the University of Hawaii's preschools,[2] and all but one had been in those preschools for at least six months. This child was the senior author's daughter; she had previously attended preschool on the mainland, and data for her fell close enough to the median of the group for the obtained measures that it was considered legitimate to retain her as an *S*, even though her father was the primary observer for the study.

Maternal interviews were not obtained for three of the children, so that the mothers of only 57 of the children provided information about maternal responses to child conflict.

RESULTS AND DISCUSSION[3]

The central problem in this study was to explore sex differences and their relations to adult-dependency behavior so as to replicate and throw additional light on a finding originally reported by Marshall and McCandless (2). Marshall and McCandless found dependency, defined grossly as total number of contacts with adults, negatively related to popularity with peers (defined by free play contacts). This negative relation was significantly stronger for girls than boys.

The present study lends confidence to this result, although the relations between gross number of adult contacts and popularity are weaker than Marshall and McCandless obtained for an Iowa population. For both sexes combined in the present Hawaiian sample, r was $-.30$ ($p \leqq .05$) when the criterion of popularity was free play association and —.16 (*ns*) when teacher judgment was used to define popularity. The correlation (using free play popularity) was higher for girls than boys (—.38, $p \leqq .05$ for girls; —.18, *ns*, for boys). While this result is in the direction of the Marshall-McCandless findings, the difference between the correlations in the present study is not significant.

---

[2] The authors appreciate the courtesy and cooperation of Mrs. Elizabeth Crooker and Miss Mary Rushit, head teachers of the intermediate and older preschool groups, in making it possible to work with their groups.

[3] Dr. Charles Truax has been extremely helpful in suggesting methods of data analysis, carrying out statistical procedures, and giving helpful critical attention to the manuscript.

In the present study dependency was broken down more refinedly than it had been by Marshall and McCandless. Dependency behaviors were separated, in accord with suggestions by Heathers (1), into *emotional* and *instrumental* dependency. Emotional dependency consists principally of seeking for comfort, affection, and support; instrumental dependency of asking objective help ("Help me put up the ladder"). However, requests for teachers to intervene in conflicts were classified by the present authors as emotional dependency. After considerable observation of behavior, the present authors agreed that the emotional, support-seeking component ("Teacher! Teacher!" or—while crying—"He hit me!") of this behavior was more significant than its instrumental aspect.

The authors hypothesized that emotional dependency accompanies (or precedes) inadequacy with peers. Excessive emotional dependency in a child may mean that he is more focused on adults than children, hence does not develop effective techniques for relating with peers. It may mean that dependency relations with parents are not yet resolved and that, accompanying this, emotionally dependent children are still sufficiently insecure that they fail to relate successfully to other children.

The hypothesis that emotional dependency would interfere with popularity was supported. When the free play criterion of popularity was used, emotional dependency and popularity correlated —.33 ($p \leqq .05$); teacher-judged popularity and emotional dependency correlated —.27 ($p \leqq .05$). It can be seen from Table 1 that girls show more total emotional dependency than boys (although the significance of this difference disappears when "asking teacher intervention" is eliminated from the category). However, rather surprisingly, "asking teacher intervention" did not interfere with popularity. The authors had expected that the child who indulged extensively in such behavior would be regarded by his peers as a "ratter" or that such behavior would interfere with the normal, competent solving of conflicts that they believed accompanied popularity. But the correlation between amount of "asking intervention" and popularity (free play) was —.08 (*ns*); with popularity (teacher judgment) it was .04.

Thus, no clear answer is provided for the question of why emotional dependency tends to interfere more with popularity for girls than boys: girls are not significantly different from boys in amount of emotional dependency when "asking intervention" is eliminated from the category; they *do* request much more intervention from teachers than boys (r_{bis} is $+.58$ for "asking intervention" and sex, with girls at the positive pole); yet asking

TABLE 1

SEX DIFFERENCES IN BEHAVIOR AS DETERMINED BY OBSERVATION*

	MEDIAN FREQUENCY OF OCCURRENCE†			
	Boys	*Girls*	χ^2	*p*
Asks teacher to solve conflict	0.6	2.2	20.17	.01
Total dependency	3.0	4.3	1.61	*ns*
Emotional dependency	1.6	3.8	5.34	.05
Instrumental dependency	1.4	0.9	0.31	*ns*
Conflicts engaged in	11.5	9.4	2.41	*ns*
Conflicts initiated	6.0	2.8	7.04	.01
Token or no resistance to conflict (%)	33.0	59.0	5.44	.02
Prolonged resistance to conflict (%)	21.7	5.2	2.41	.20
Persists in activity after conflict (%)	98.0	82.0	8.42	.01
Changes activity after conflict (%)	9.2	26.2	7.04	.01
Contacts with other children	51	48	0.34	*ns*
Contacts with adults	9.2	8.7	0.12	*ns*
Instrumental nurturance offered by adults	3.0	3.0	0	*ns*
Emotional nurturance offered by adults	0.8	0.9	0.21	*ns*
Teacher popularity ranking (weighted score)	50	48	0.26	*ns*

* The Yates correction for one degree of freedom has been used in computing chi square.

† Legends indicate when percentages or weighted scores were used.

intervention is not related to popularity, while emotional dependency is significantly related in a negative direction, and more strongly so for girls than boys.

The most plausible speculation that occurs to the authors concerns cultural differences in treatment of emotional dependency between girls and boys: it is moderately acceptable for girls, relatively unacceptable for boys. Girls, then, who have difficulty in peer relations are likely to "run to the teacher," but boys with similar problems (who have received discouragement for emotional dependency) seek other methods of compensating, perhaps in some cases by ruggedly independent or aggressive behavior, so that what relation there might be for them between emotional dependency and popularity is attenuated and possibly in some cases negated (i.e., the culture forces "compensation for emotional dependency" on boys, but not on girls).

It was predicted that boys and girls would not differ in instrumental or total dependency. Table 1 shows that this prediction was confirmed. Neither was it expected that instrumental dependency would be related to popularity. This proved to be the case: free play popularity and instrumental dependency correlated —.14; popularity as judged by teachers and instrumental dependency correlated —.08.

Another hypothesis was that, in addition to finding that "asking intervention" from teachers would be more frequent for girls than boys, the maternal interviews would show that girls had learned such behavior from their mothers' methods of handling their conflicts: i.e., mothers would let boys "go it alone" more frequently, while they would set themselves up more often as problem solvers for their daughters. But biserial correlations with these two maternal "scales" were only —.09 and —.15, respectively. This lack of relation is most likely due to the low reliability of categorizing the interview, lack of validity in the interview, the limited sampling of mother-child interactions, or a combination of these factors.

Table 1 supplies certain other data of interest, although bearing less directly on the major purposes of the study: Boys, significantly more often than girls, initiate conflicts. That boys are more task-oriented than girls is indicated by the greater frequency with which they persisted in the activity in which they were engaged when conflict occurred and the lesser frequency with which they changed to another activity following conflict. Girls more often than boys offer little or no resistance to conflict. The sexes do not differ in popularity as judged either by teachers or from free play, nor in either instrumental or emotional nurturance offered by adults, nor in number of contacts with adults (a rough index of total dependency). It is the *nature* of these adult contacts that seems to make the difference.

SUMMARY

The central problem of this study was to investigate "refinements" of dependency behavior as they vary by sex and affect children's relations with their peers. The authors were also interested in the relation between "general" adult-dependency and popularity. A general postulation was that girls would show more emotional dependency than boys and that emotional—but not instrumental—dependency on adults would interfere with social relations with peers.

Twenty-six Caucasian (*haole*) and 34 non-Caucasian (*nonhaole*) University of Hawaii preschool children were used as subjects; 23 were boys, 32 girls.

Popularity and emotional dependency were found to be negatively related, as were "total adult contacts" and popularity. This over-all index of adult dependency interfered more with girls' than with boys' popularity, although the difference between sexes was not as great as had been found in an earlier study done with University of Iowa preschoolers.

Other findings indicated that mothers did not intervene more frequently in daughters' than in sons' conflicts, although girls more frequently than boys asked teachers to solve their conflicts. Girls showed more total emotional dependency than boys, although the difference became nonsignificant when "asks teacher intervention" was eliminated from the category of emotional dependence. Girls were less likely to resist conflicts and were

more likely to change activity following conflict than boys. Girls initiated fewer conflicts, but engaged in almost as many as boys. There were no sex differences in instrumental or total dependency, and instrumental dependency did not interfere with popularity. There were no sex differences in total or in instrumental dependency or in popularity, whether this was judged by teachers or measured by associations during free play.

REFERENCES

1. HEATHERS, G. Emotional dependence and independence in nursery school play. *J. genet. Psychol.*, 1955, 87, 37-57.
2. MARSHALL, H. R., & MCCANDLESS, B. R. Relationships between dependence on adults and social acceptance by peers. *Child Develpm.*, 1957, 28, 413-419.

SEX OF EXPERIMENTER AND SEX-TYPED BEHAVIOR OF YOUNG CHILDREN

L. J. Borstelmann*[1]

Stanford University

Although in recent years considerable attention has been given to the study of parents as significant contributors to child behavior patterns and tendencies (3, 8, 9, 12, 13, 15, 16), there has been little evident awareness of experimenters as adult interactors varying in attributes or behaviors that may affect responses elicited from children. Whereas variations in parental attitudes and behaviors are often hypothesized as critical antecedents, experimenters seem to be regarded as generalized adult figures, impersonal behavior initiators, and data recording machines. If the child is indeed responsive to such qualities in his parents as warmth, permissiveness or restriction, sex roles, etc., then he should build up certain expectations and anticipatory sets associated with the occurrence of such qualities in other significant adults. Yet studies designed to manipulate child behavior have generally failed to consider or control for relevant experimenter characteristics. Some attention has been directed to the presence or absence of adults as a factor affecting child behavior (14), but the effects of *E* attributes or behaviors have been largely neglected.

One of the more obvious, readily defined *E* attributes that might be considered to influence a child's responses is sex status. Interactions with male versus female experimenters may be expected to involve some differences in response sets of the experimental children. A few recent studies have varied sex of *E* as part of the experimental design. Gewirtz and his colleagues (6) in two studies of attention seeking with variation of adult

---

* Department of Psychology, Duke University, Durham, North Carolina.

[1] This study was conducted while the author was a Research Fellow of the United States Public Health Service. Appreciation for advice, guidance, and cooperation is acknowledged to Robert R. Sears, Director, Laboratory of Human Development, and Edith Dowley, Director, Nursery School, Stanford University, and their staffs.

Child Develpm., 1961, 32, 519-524.

availability found a significant cross-sex pattern of child response, both boys and girls making more dependency bids to the opposite sex *E*. Rosenblith (11) found a male *E* to be more effective than a female *E* in eliciting imitation learning. Child (2) found a greater frequency of distant versus immediate goal choices with a female *E*. All of these investigators employed only one *E* of each sex and are properly cautious in noting that the obtained effects might be due to *E* characteristics other than sex difference. Such equivocal conclusions represent the usual state of affairs when the issue is peripheral to the primary interests of the investigators. We are left either with an unwarranted generalization from an *N* of 1 or with the tantalizing suggestions of an important effect that is unfortunately not verified by representative sampling of the presumably relevant variables. Hammond (7), following Brunswik's notion of representative versus systematic design, has pointed out that unwarranted conclusions concerning the effects of examiners occur as a result of failure to establish the crucial independent variable. The present study was designed to implement this *E* sampling requirement with respect to the effects of sex of *E* upon the sex-typed behavior of young children.

In considering the possible effects of *E* sex status upon the degree of the child's appropriately sex-typed responses, two contradictory hypotheses are suggested: (a) that an *E* of the same sex as the child will enhance sex-appropriate responses of the child by providing reinforcement as an identification model; or (b) that an *E* of the opposite sex to the child will enhance sex-appropriate responses by providing reinforcement for assertion of the complementary role. The possibility that either or both hypothesized effects might operate within a given group of children led to our investigation of sex of *E* as a variable in a test-retest design. The alternative possibilities of *E* influence will be evident as follows:

1. If the male versus female status of *E*s produces a constant effect among boys or girls, this will be reflected in different degrees of sex-appropriate behavior according to sex of *E*.

2. If sex of *E* produces a variable effect among children, then test-retest agreement will be greater when sex of *E* is unchanged than when sex of *E* is reversed.

PROCEDURE AND SUBJECTS

The study design involved initial and repeat administration of three experimental measures of sex-typed behavior under conditions of *E* sex constancy and of *E* sex reversal. Each of eight male and eight female *E*s administered all three measures to two boys and two girls. All children had a change of *E* for readministration, but half of the subjects of each sex had a new *E* of the same sex as the initial *E* and half had a shift in sex of *E*. Thus, the 32 boys and 32 girls were assigned randomly to four sequences of *E*s: male-male, male-female, female-female, and female-male.

The experimental tasks were three measures of sex-typed behavior used by Sears and his group in their current study of identification in young children: the It Scale developed by Brown (1) and two instruments especially constructed for the project, Toy Preference (TP) and the Pictures Test (Pix). TP was adapted from a procedure of Rabban (10), using seven male and seven female toys with the child instructed to make eight consecutive choices of preferred items. This technique has proven to have consistently high interest value for nursery school children. The Pictures Test, adapted from a method used by Fauls and Smith (5), is a set of 12 items with separate forms for each sex, wherein the child is depicted in combinations of male or female, adult or child activities in the presence of both parent figures. The subject is asked to make a simple choice of preferred activity in a series of paired comparisons. This technique provides a measure of adult role preference as well as sex role preference.

The 64 child *S*s were all enrolled in the University Nursery School, where the study was conducted. They ranged in age from 3 years, 4 months to 5 years, 0 months with a median age of 4-2 and no age difference between boys and girls. The 16 *E*s were university students in child study courses, familiar with the nursery school. All *E*s were pretrained in the use of the experimental measures.

RESULTS

The first hypothesis regarding any general effects of sex of *E* upon degree of children's sex-typed responses was examined in terms of differences between male and female *E*s on initial administration of the three measures and differences between the two readministration conditions of constancy versus change in sex of *E*. There were no significant mean differences due to sex of *E* for boys or girls on the initial measures. There were no significant mean differences between test 1 and test 2 for boys or girls regardless of *E* sequence. Thus, there is no evidence that sex of *E* produces any constant effect of either increasing or decreasing the sex-appropriate responses of boys or girls. Boys had a consistently and significantly higher mean sex-appropriate score on the Toy Preference and It Scale, but not on the Pictures Test.[2]

The second hypothesis of variable effects of sex of *E* upon child response was examined in terms of any differences between retest reliabilities for the experimental conditions of constancy versus reversal of sex of *E*. If there is interaction between sex of *E* and identity of *S* (as distinct from that between sex of *E* and sex of *S*), then the reliabilities will be lower for male-female and female-male sequences than for male-male and female-female. As may be seen from the results as presented in Table 1, there are no significant differences among the Pearson correlations for any of the measures.

---

[2] Information about the sex-typed measures, scoring criteria, and distributions will be provided upon request to the Laboratory of Human Development, Stanford University.

TABLE 1

TEST-RETEST RELIABILITY COEFFICIENTS FOR THREE SEX-TYPING MEASURES WITH STABILITY AND REVERSAL OF SEX OF EXPERIMENTER (*E*)

Experimental Condition	*N*	TP		It		Pix	
		B	G	B	G	B	G
E Sex Constant	16	.60§	.56§	.79†	.63‡	.38	.19
E Sex Reversed	16	.73†	.74†	.81†	.66‡	.38	.52§
Total	32	.76*	.62*	.80*	.64*	.37§	.38§

* $p < .001$. † $p < .01$. ‡ $p < .02$. § $p < .05$.

Again there is no evidence that sex of *E* has any significant effect on the sex-typed choices of the subjects. The reliabilities actually tend to be slightly higher with *shift* in sex of *E* for both sexes on Toy Preference and for girls on the Pictures Test.

Two further aspects of the data were examined: age effects and agreement among measures. Since age of *S*s is distributed randomly among *E* sequences, any age differences in sex-appropriate responses may be ascertained. For this purpose both boys and girls were separated into older and younger groups in accordance with the median split of 4 years, 2 months. Older boys showed a general trend of more sex-appropriate responses, but the difference was significant only for the It Scale. There were no age differences in retest consistency for boys. Girls reflected consistent age differences only on the It Scale, with older girls more sex-typed and younger girls less reliable.

Finally, the intercorrelations among the three measures as shown in Table 2 are all positive but not consistently significant. TP and Pix are significantly correlated for both sexes on initial measures, but the significance

TABLE 2

INTERCORRELATIONS AMONG THE SEX-TYPING MEASURES

	BOYS ($N=32$)		GIRLS ($N=32$)	
	Test 1	*Test 2*	*Test 1*	*Test 2*
TP – It	.33	.35*	.25	.30
TP – Pix	.44*	.33	.51†	.09
It – Pix	.26	.40*	.28	.39*

* $p < .05$. † $p < .01$.

drops out on the retest measures. It and Pix are not significantly related on test 1 but become so on test 2 for both boys and girls.

DISCUSSION

Male or female status as an attribute of *E*s has been investigated for possible effects upon sex-appropriate behavior of young boys and girls with negative results. Thus, the sex role aspects of the adult-child interaction must be considered negligible with respect to the experimental measure of sex-typed behaviors. This conclusion does not suggest that Gewirtz' and Rosenblith's findings of differences related to sex of *E* are fortuitous, since they were studying behaviors of dependency and imitation that were demonstrably a function of *E* behavior. In the present study the tasks employed were intrinsically interesting to the children, and the *E*s were permissive with no cues to reinforce certain responses. Thus, the situation was structured to maximize the child's recourse to his own personal preferences. There were no known external referents that might direct responses, other than sex of *E*, and this was apparently irrelevant to the children. There of course remains the possibility that some differential effects due to *E*s did occur, but we have sampled only the dimension of sex status and our sample size does not permit analysis of individual *E* effects.

The negative results of the present study do underscore the necessity of adequate sampling of *E*s for determination of any experimental effects attributed to *E* differences. As Gewirtz and Rosenblith both note, their differences between one male and one female *E* may well be due to many individual characteristics other than their obvious sex status. In the present instance we have systematically studied sex of *E* as a relevant experimental variable and have found no effects evident. We do not conclude therefrom that children are indifferent to sex of adults but rather that any inferences about the effects of *E* attributes or behaviors upon child response can only be justified on the basis of an adequate sampling of *E*s on the relevant characteristics with reference to given behavior systems of the child.

The present findings confirm the reliability estimates reported by Brown for the It Scale and by Rabban for a similar toy preference technique. The low reliabilities of the Pictures Test are in line with other uses of paired comparisons procedures with nursery school children, as reported by Clifford (4).

Since the intercorrelations among the three measures employed are quite low, there remains some question about the generality of a concept of sex-typing at the nursery school age, at least insofar as toy and play preferences are concerned.

SUMMARY

Alternative hypotheses of constant or inconsistent effects of experimenter sex difference upon the responses of young children to three measures of

sex-typing were studied by means of a test-retest design under two conditions of change or constancy of sex of experimenters. The results yielded clear evidence that sex of experimenter was not a significant variable. Sex and age differences in degree of appropriate sex-typing, and the limited communality among the measures, were noted.

References

1. Brown, D. G. Sex-role preference in young children. *Psychol. Monogr.*, 1956, 70, No. 14.
2. Child, I. L. Children's preference for goals easy or difficult to obtain. *Psychol. Monogr.*, 1946, 60, No. 4.
3. Child, I. L. Socialization. In G. Lindzey (Ed.), *Handbook of social psychology*. Addison-Wesley, 1954. Pp. 655-692.
4. Clifford, E. Ordering of phenomenon in a paired comparisons procedure. *Child Develpm.*, 1959, 30, 381-388.
5. Fauls, L. B., & Smith, W. D. Sex-role learning of five-year-olds. *J. genet. Psychol.*, 1956, 89, 105-119.
6. Gewirtz, J. L., Baer, D. M., & Roth, C. H. A note on the similar effects of low social availability of an adult and brief social deprivation on young children's behavior. *Child Develpm.*, 1958, 29, 149-152.
7. Hammond, K. R. Representative vs. systematic design in clinical psychology. *Psychol. Bull.*, 1954, 51, 150-159.
8. Levy, D. M. *Maternal overprotection*. Columbia Univer., 1943.
9. Miller, E. R., & Swanson, G. E. *Inner conflict and defense*. Holt, 1960.
10. Rabban, M. Sex-role identification in young children in two diverse social groups. *Gent. Psychol. Monogr.*, 1950, 42, 81-158.
11. Rosenblith, J. Learning by imitation in kindergarten children. *Child Develpm.*, 1959, 30, 69-80.
12. Sears, R. R., Maccoby, E. E., & Levin, H. *Patterns of child rearing*. Row, Peterson, 1957.
13. Sears, R. R., Whiting, J. W. M., Nowlis, V., & Sears, P. S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, 47, 135-236.
14. Siegel, A., & Kohn, L. Permissiveness, permission and aggression: the effect of adult presence or absence on aggression in children's play. *Child Develpm.*, 1959, 30, 119-130.
15. Watson, R. I. *Psychology of the child*. Wiley, 1959.
16. Whiting, J. W. M., & Child, I. L. *Child training and personality*. Yale Univer., 1953.

MATERNAL ACCEPTANCE OF RETARDED CHILDREN: A QUESTIONNAIRE STUDY OF ATTITUDES AND RELIGIOUS BACKGROUND

G. H. Zuk,* Ralph L. Miller, John B. Bartram,
and Frederick Kling[1]

St. Christopher's Hospital for Children, Philadelphia

This study proposes to examine the relation of religious background to maternal acceptance of children. In this case, the children are retarded. Maternal acceptance will be explored through means of a questionnaire tapping attitudes of the mothers toward their children.

Guilt as an important psychological process in parent-child relations has recently been examined in a study by one of the writers (27). Catholic mothers were judged to be more acceptant of their retarded children than non-Catholic mothers. This result was interpreted in the context of the Catholic dictum that parents not blame themselves for bearing a retarded child. Many writers have isolated guilt as a major factor impeding parental acceptance of handicapped children. Sarason (16), Scheimo (19), Waterman (24), Rheingold (12), Hastings (6), and Stone (21) have made this point in reviewing problems in acceptance of retarded children. As it affects the acceptance of cerebral palsied children, it has been mentioned by Perlstein (10), Wishik (25), Moore (9), and Roe (14).

Two recent studies seem to support the finding that parental acceptance of retarded children is greater among Catholics than non-Catholics. One by Saenger (15) reports a significantly higher percentage of institutionalization among Jews than Italians, a predominantly Catholic group. In another, Farber (5) notes that the marital integration of Catholic families is less disturbed by the presence of a retarded child in the home; he concludes that ". . . non-Catholic families seem more affected than Catholics in a crisis situation" (p. 65).

In their book on the Hutterites, Eaton and Weil (4) point to the high level of acceptance of the retarded in the community. In this sect, the retarded

---

* St. Christopher's Hospital for Children, 2600 North Lawrence Street, Philadelphia 33, Pennsylvania.

[1] The first three authors are, respectively, psychologist, chaplain, and pediatrician (director) of the Mental Retardation Clinic at St. Christopher's Hospital for Children. The fourth author acted as statistical consultant on the study. The study was supported in part by grants from the following agencies: Children's Bureau, Department of Health, Education, and Welfare; Pennsylvania State Department of Health, Division of Child and Maternal Health; and the Dollfinger-McMahon Foundation. The material presented in this study also served as the basis for an unpublished master's thesis (8).

Child Develpm., 1961, 32, 525-540.

child is not held personally responsible for the mischief he may commit because he is thought not to have reached the age of accountability before God. There seems to be a parallel between the Hutterite and Catholic viewpoints, as described for the Catholics, for example, in a religious tract by Breitenbeck (3).

Some apparently contradictory evidence comes from a recent study by Boles (2). Although his subjects were mothers of cerebral palsied children, this group of children overlaps the retarded greatly in mental ability. Boles analyzed the responses to an extensive personality questionnaire and found that Catholic mothers verbalized more guilt and anxiety and were more socially withdrawn than non-Catholic mothers. This finding applied to Catholic mothers of nonhandicapped children as well, so its interpretation in the light of the handicap variable is ambiguous. However, one may interpret it as evidence for lesser acceptance of their children by Catholic mothers.

This study hopes to provide a step toward clarification of the relation between a cultural factor, religious background, and a psychological concept, maternal acceptance. It hopes to provide evidence bearing on the note, often made in the clinical literature, that guilt impedes parental acceptance of handicapped children. It hopes to provide evidence also bearing on another point made in the clinical literature, that attitudes expressed by parents of handicapped children about these children often may not accurately reflect actual practices.

PROCEDURE AND METHOD

The Questionnaire

A questionnaire was devised to elicit information about (a) the religious practices and (b) attitudes, feelings, and beliefs of mothers of retarded children about retarded children.[2] Part 1 of the questionnaire was composed of two sections: section A (14 items) inquired about the mother's religious practices; section B (10 items) inquired of the mother her husband's religious practices. However, something more than the concrete practice of religion was involved in responding to the items of part 1. Some of the items attempted to explore the individual's orientation toward religion, i.e., the intensity of faith. The following items should serve as examples of what was implied: "My faith has been a source of personal strength to me in everyday living"; "I have received real help with my problems from the other members of my religious faith"; "My faith has been strengthened as a result of having a retarded child."

Part 2 of the questionnaire was composed of 50 items reflecting attitudes of mothers about their retarded children. A number of these items were borrowed from the Parental Attitude Research Instrument (18), but revised according to what were believed to be requirements of the sample. The PARI was considered too long and to require too high a level of verbal sophistication. In addition, the relevancy of many of the items of the PARI for the particular problem under study was questioned. Items were worded specifically to orient mothers to revealing attitudes about their retarded children, not about their other children or children in general. They could respond in one of four ways: strongly agree, mildly agree, mildly disagree, strongly disagree. A selected sample of the 50 items is as follows:

---

[2] A limited number of copies of the questionnaire are available and may be had upon request of the authors.

1. I frequently ask: Why did my child have to be slow?
2. I think an operation would cure or reduce my child's problem.
3. Frequently I give my slow child more love than he needs.
4. It is not my fault that our child is retarded.
5. My greatest mission in life is to care for my slow child.
6. I have not hesitated in telling our relatives what is wrong with our child.
7. Often my husband and I disagree on how to handle our child.
8. My husband and I are making plans for our child's future care and training.
9. I frequently become very impatient with my slow child.
10. In some ways my child is superior to normal children.
11. I do not discipline my child enough.
12. One of the most difficult things about taking care of a slow child is a woman feels that she can't get out.
13. I feel that my child was normal at birth but started having problems after he (or she) was older.
14. Having to care for a slow child has solved some of my more difficult problems.
15. I occasionally feel that something I have done has caused my child to be slow.
16. It sometimes bothers me to think of sending my child to an institution for slow children.
17. When I see all the normal children in the world, I frequently get disturbed that my child is not like the others.
18. God has chosen me to take care of a retarded child.
19. I tend to protect my child too much.
20. Even when I feel like showing my love for our slow child, I often have trouble.

This sample of items should make clear the effort to probe for and elicit feelings of guilt, anxiety and rejection on the part of mothers.

Subjects

The questionnaire was mailed to 125 mothers who had had contact with the Mental Retardation Clinic at St. Christopher's, a nondenominational hospital for children, from about September, 1957, to December, 1958. A letter describing the purpose of the questionnaire and the need for early return accompanied it.[3] As a result of numerous follow-ups by mail and telephone, 75 questionnaires were returned, 60 per cent of the maximal expected. Several reasons for failure to return the questionnaire were later discovered: lack of ability to follow directions (a few mothers were foreign-born and could not read English sufficiently well), change of address, and refusal to answer items for personal reasons. For these and other reasons, returns of the order of 60 per cent are typical in cases of such self-selected samples as the present. Considerably more elaborate studies in recent years have hardly exceeded this order of return; for example, *see* (20).

---

[3] The covering letter was sent out under the signature of one of the authors, Rev. Ralph L. Miller, a minister of the Presbyterian denomination.

The group returning the questionnaire did meet the methodological requirement of being composed of about equal numbers of Catholics and non-Catholics. In addition, it was felt to be fairly homogeneous for education and socioeconomic status.

Table 1 gives some descriptive information about the mothers (*N* of 72, because the questionnaires of three mothers failed to supply needed identification material). There were 37 Roman Catholic mothers and 35 non-Catholic mothers. Of the non-Catholic mothers, 27 were Protestant and eight Jewish. Of the 27 Protestant mothers, 24 were white and three Negro. Information is given in Table 1 for numbers of children in the sample and sex of the child. Average formal education of the mothers was 11 years. The majority of fathers were employed in skilled and semiskilled trade.

TABLE 1

DESCRIPTIVE INFORMATION ON MOTHERS AND CHILDREN

	CATHOLIC ($N=37$)		PROTESTANT ($N=27$)		JEWISH ($N=8$)	
Variable	*Mean*	*Range*	*Mean*	*Range*	*Mean*	*Range*
Age of mothers	33.5	24–46	32.0	24–51	37.2	29–46
Age of children	4.7	1–12	5.0	1–12	6.0	2–13
Number of siblings	2.5	0– 7	2.0	0– 6	1.6	1– 3
Most recent IQ of children	44.9	12–75	48.3	22–85	34.8	23–50

Scoring the Questionnaire

On part 1 of the questionnaire, responses were weighted as reflecting greater or lesser formalized religious observance. For example, the response "weekly" to the item, "I attend worship services . . ." received a weight of 5, the response "monthly" a weight of 3, the response "yearly" a weight of 1, and the response "not at all" a weight of 0. Several of the items were combined to yield a composite weighted score. For these, simple scoring matrices were developed.

On part 2 of the questionnaire, an attitude scale, the individual (R.L.M., a Protestant minister) who constructed the scale also prejudged responses as reflecting greater or lesser acceptance of the child by the mother. An effort to determine to what extent staff members agreed with the prejudged responses was carried out by asking six members of the Clinic staff to fill out the questionnaire as they imagined an ideally acceptant mother might. Agreement of three or more of the staff with the prejudged rating occurred in 47 of the 50 items. On three remaining items, two of the staff agreed with the prejudged rating. None of these three items, however, on which agreement between staff and judge fell below 50 per cent was found to evoke differential response among the religious groups. While 50 per cent agreement between staff and judge can hardly be called "high," it seemed the best procedure to be as arbitrary as this in view of the state of knowledge of what constitutes maternal acceptance as reflected in attitude expression. Higher agreement concerning acceptance might have been found, but it seems more likely as a result of arbitrary compromise than obvious decision.

The 50 items of part 2 were broken down into five subscales, according as their content seemed to reflect the following factors: Overprotection (8 items), Discipline (10 items), Cause of Retardation (8 items), Acceptance of Diagnosis (9 items), and Mother's Self-Fulfillment (15 items).

Thus, from parts 1 and 2 of the questionnaire a total of eight scores were obtained: one of mother's religious practices, one of father's religious practices, five from subscales of part 2, and one from the total scale of part 2. In addition, each item of the full questionnaire was statistically evaluated to try to detect religious differences in response.

Clinical Measures

Clinical judgments of acceptance of the child by mother and adjustment of the child were obtained from the staff social worker and pediatrician.[4] Acceptance ratings were based on absence of anxiety in the mother, feelings of adequacy with the child, and absence of hostility toward child or professional guidance. The impressions of the social worker were used primarily because: (a) the focus of her interviews with parents was more directly on their reaction to the child, and (b) she was a member of neither of the predominant religious groups compared in the study. Reported in a previous study (27) are levels of agreement between this worker's judgments and those of two psychologists who independently rated samples of social history protocols on which all judgments were based. In view of what was considered to be sufficiently high agreement among these individuals, it was deemed appropriate to use the social worker's judgments as a clinical measure of acceptance.

For 35 of the 72 mothers, the social worker's judgments were already available from the previous study. These evaluations had been made at the time of initial contact with the mother. For the remaining 37 mothers, however, the social worker was asked to make her ratings as of the time of the present study. The set of 72 ratings was correlated with the other variables of the study (*see* Table 5). The set of 35 ratings made at initial contact was also correlated with the other variables of the study, but failed to show results differing significantly from those using the larger set of 72.

---

[4] Special thanks are due Miss Helen Beck, social worker, and Dr. June Dobbs, pediatrician, for their evaluations of the mothers.

Table 2 lists, by religious background, the social worker's ratings of 35 mothers made at the time of initial contact and her 37 ratings made later. These ratings were then combined (N of 72) and compared with acceptance ratings made for the previous Zuk study. In the previous study, a significant correlation (chi square equalled 18.28, corrected for continuity; $p < .001$) was established between religious background and acceptance. In the present study, while a trend in the same direction was apparent, there was no significant correlation (chi square equalled 1.44, corrected for continuity; $p > .20$). The social worker rated as acceptant 19, or 51 per cent, of 37 Catholic mothers; 10, or 37 per cent, of 27 Protestant mothers; and two, or 25 per cent, of eight Jewish mothers.

Inspection of the 35 ratings carried over to the present study suggests the possibility of some bias in the return of the questionnaires. There was a comparatively higher percentage of return from mothers judged by the social worker to be nonacceptant. In the previous study, the ratio (based on an N of 39) of acceptant to nonacceptant Catholics was approximately 2:1. But in the 35 ratings carried over to the present study (i.e., mothers who returned the questionnaire), the ratio (based on an N of 18) was 1:1. In the previous study, the ratio (based on an N of 28) of acceptant to nonacceptant Protestants was about 1:5. But in the 35 ratings carried over, the ratio (based on an N of 13) was about 1:2.

Significant religious differences were not apparent in the pediatrician's ratings of the child's adjustment, although it is of some interest that there was a slight tendency to rate Catholic children toward the lower end of the scale. Of the Catholic children, one was rated excellent in adjustment, three

TABLE 2

SOCIAL WORKER'S EVALUATIONS OF MATERNAL ACCEPTANCE

	EVALUATIONS							
	"Old" ($N=35$)		*"New"* ($N=37$)		*Present Study* ($N=72$)		*Zuk Study* ($N=76$)	
	A*	NA*	A	NA	A	NA	A	NA
Catholic Mothers	9	9	10	9	19	18	25	14
Protestant Mothers	4	9	6	8	10	17	5	23
Jewish Mothers	0	4	2	2	2	6	0	9

NOTE.—Included for comparison are social worker's "old" (carried to present from a previous study) and "new" (obtained for present study) evaluations of mothers' acceptance of child, and, also, evaluations made on mothers ($N=72$) who returned the questionnaire (present study, "old" plus "new") and on mothers ($N=76$) who composed sample of prior Zuk study (27).

* A = Acceptant; NA = Nonacceptant.

good, 21 fair, and 12 poor. Of the non-Catholic children, two were rated excellent in adjustment, five good, 21 fair, and seven poor.

The lack of significant correlation ($r = .21$; *see* Table 5) between the pediatrician's and social worker's judgments suggests that nonoverlapping criteria were being utilized for the judgments.

The Factor Analysis

A factor analysis (*see* Table 6) of the variables employed in the study was carried out under direction of one of the authors (F. K.). A modification of the group method approximating principal components was used (22). Special care was taken to limit the weighting of the dichotomous religious variables so as not to allow them to overdetermine the structure, since our primary interest was in the structure emerging from the acceptance and religious involvement variables.

The table of intercorrelations (*see* Table 5) revealed certain clusters of variables which tended to be positively correlated. Table 5 was reduced to its major components by the method of factor analysis described above. The factors seemed sufficient to account satisfactorily for the observed intercorrelations. The factor loadings (orthogonally rotated) of each variable are given in Table 6.

In the present case, it was inferred that factor I was characterized by acceptance of the retarded child and factor II by religious involvement. These factors were found to be more independent than anticipated. However, it was found that certain aspects of acceptance (namely, the variables, Overprotection and Discipline) tended to be related to religious involvement.

RESULTS

Split-half reliability coefficients of sections A and B of part 1 were .91 and .92, respectively. Table 3 shows mean scores obtained from sections A and B. Catholic mothers reported themselves and their husbands as more intensive in formalized religious observance than Protestant or Jewish mothers. Significant religious differences[5] were found in this direction on three of the 24 items on part 1, with differences on a fourth item approaching the 5 per cent level. Catholic mothers, to a greater degree than Protestant mothers, reported that they were:

1. More faithful in church attendance;
2. More loyal to their parents' religious training;
3. More prone to attend church when everything was going well, rather than just in time of crisis;
4. More consistent in their practice of prayer.

Split-half reliability coefficients of the subscales of part 2 were .58 (Overprotection), .54 (Discipline), .72 (Cause of Retardation), .67 (Acceptance of Diagnosis), and .64 (Mother's Self-Fulfillment). Table 3 also shows mean scores obtained on these subscales. Clear differences among the religious groups are not apparent, although the sums of mean subscale scores do reflect differences in the predicted direction. For the Catholic group, the sum is 26.2; for the Protestant group, 23.9; and for the Jewish group, 24.2.

---

[5] The p levels were determined by the formula given in Walker and Lev (23, pp. 77-78).

TABLE 3

AVERAGE SCORES OF MOTHERS, BY RELIGIOUS GROUP, ON QUESTIONNAIRE

	Catholic ($N=37$)	*Protestant* ($N=27$)	*Jewish* ($N=8$)	*Maximum Score Possible*
Part 1 (Sections A & B):				
Mother's Religious Observance	45.0	39.0	20.0	58
Father's Religious Observance	24.0	15.0	11.0	34
Sums	69.0	54.0	31.0	92
Part 2 (Subscales):				
Mother's Self-Fulfillment	7.2	7.4	7.0	15
Discipline	6.0	5.4	5.8	10
Overprotection	3.0	2.4	1.6	8
Acceptance of Diagnosis	5.7	4.8	6.5	9
Acceptance of Cause	4.3	3.9	3.3	8
Sums	26.2	23.9	24.2	50

Table 4 reveals some measure of association between parts 1 and 2 of the questionnaire. Taken by religious groups, individuals scoring "high" and "low" on part 1 are compared as to their scores on the subscales of part 2. "Highs" and "lows" of the Catholic and Protestant group were split at the upper and lower 27 per cent level. The Jewish group, because of its small size, was split 50-50. Significant differences within individual subscales are not apparent, but on the subscales taken together significant differences in the predicted direction occur: "high" Catholics scored higher on all subscales than "low" Catholics ($p < .03$), and "high" Catholics scored higher on all subscales than "high" Protestants ($p < .03$).

Significant religious differences were found on only three items of part 2. On each of these, however, Catholic mothers more frequently gave the prejudged response reflecting greater acceptance of the child. More frequently than Catholic mothers, Protestant mothers stated that:

1. Their greatest mission in life was to care for their retarded child;
2. They more often became impatient with their retarded child;
3. Medical care could bring their child back to normal.

Taking all 50 items of part 2 together, the average percentage of "acceptant" response of the Catholic group was 26.5 per item (with a range of 2 to 91 per cent). The average percentage of "nonacceptant" response by the Catholic group was 21.6 per item (with a range of 0 to 50 per cent). The average percentage of "acceptant" response by the Protestant group was

TABLE 4

COMPARISONS OF SCORES ON PART 2 OF QUESTIONNAIRE OF "HIGH" AND "LOW" SCORERS ON PART 1, BY RELIGIOUS GROUP

	CATHOLIC		PROTESTANT		JEWISH	
Variables	*"Highs"* ($N=10$)	*"Lows"* ($N=10$)	*"Highs"* ($N=8$)	*"Lows"* ($N=8$)	*"Highs"* ($N=4$)	*"Lows"* ($N=4$)
Part 2 (Subscales):						
Mother's Self-Fulfillment	7.9	7.1	7.6	7.3	7.7	4.8
Discipline	6.7	5.9	6.2	5.3	6.7	4.8
Overprotection	4.1	2.6	3.1	2.0	1.1	2.0
Acceptance of Diagnosis	6.3	5.4	5.0	5.0	5.2	7.8
Acceptance of Cause	4.8	3.5	4.4	4.4	3.2	3.5
Sums	29.8	24.5	26.3	24.0	23.9	22.9

24.1 per item (with a range of 4 to 82 per cent). The average percentage of "nonacceptant" response by the Protestant group was 23.5 per item (with a range of 0 to 63 per cent). The differences are slight but in the predicted direction. It is of interest also that the average percentage of omitted responses per item was somewhat greater for the Protestant group (4.8 per cent with a range of 0 to 19 per cent) as compared with the Catholic group (4.1 per cent with a range of 0 to 27 per cent).

Some of the items composing part 2 approximate the wording of items that have been used in other studies of parents of handicapped children. For example, Ray (11) reported that 70 per cent of his sample of mothers of cerebral palsied children replied in the affirmative to the question: "Do you think God has given you a handicapped child as a 'cross to bear'?" In the present study, 61 per cent of the Catholic mothers (based on an *N* of 37), 29 per cent of the Protestant mothers (based on an *N* of 27), and 25 per cent of the Jewish mothers (based on an *N* of 8) replied in the affirmative to the item: "Having a slow child is the cross I must bear." Nineteen per cent of the Protestant mothers made no response to this item, but only 3 per cent of the Catholic mothers did not respond.

Boles (2) found that 68 per cent of his sample of mothers of cerebral palsied children and 70 per cent of his mothers of nonhandicapped children answered in the affirmative to the item: "I believe that if one puts in enough effort any condition can be changed." To an item in the present study, "I think my child will outgrow his problem," 45 per cent of the Catholic mothers, 52 per cent of the Protestant mothers, and 24 per cent of the Jewish mothers replied in the affirmative.

The importance of phrasing of items is highlighted by contrasting the responses to the item, "Having a slow child is the cross I must bear," (*see* above) with those to the item, "My greatest mission in life is to care for my slow child." On the latter, 89 per cent of the Protestant mothers, 61 per cent of the Catholic mothers, and 62 per cent of the Jewish mothers agreed. Although both items on the surface appear to be tapping the same attitude, a religious difference seems to have sprung up. One may conjecture that the Catholic group responded more positively to the phrase "cross to bear," the Protestant group to the phrase "mission in life."

Table 5 presents the results of a Pearson correlational analysis of the variables employed in the study. The correlation of section A with section B of part 1 was .48. The intercorrelations among the subscales of part 2 ranged from .23 (Overprotection with Acceptance of the Diagnosis) to .50 (Acceptance of Cause with Acceptance of Diagnosis). The correlations of scores reflecting intensity of mothers' religious observance with religious background varied from —.60 (Jewish) to —.05 (Protestant) to .42 Catholic). The range of correlations between intensity of fathers' religious observance and religious background was more restricted, —.21 to .27, but the directions were the same as those for mothers. Clinical judgments of the mother or child failed to correlate significantly with scores from parts 1 or 2. Correlations between the social worker's ratings of mothers' acceptance and parts 1 and 2 ranged from .05 to .22. Correlations between the pediatrician's ratings of adjustment of the child and parts 1 and 2 ranged from —.13 to .20.

As might be expected, the correlation between mothers' and children's ages was significant but slight, .28. A significant correlation (.34) was found between IQ and the pediatrician's ratings, which reflected a reasonable kind of bias in these ratings. It is rather more difficult to make sense of some of the other correlations that reached or exceeded the .05 level of significance: for example, the correlations between intensity of mother's religious practices and IQ of the child (.27), between the Acceptance of Diagnosis subscale and sex of child (—.27), or between the Discipline subscale and age of child (—.24), are hard to interpret. They must remain unexplained for the present or perhaps to be explained as chance occurrences.

Table 6 reports the results of a factor analysis of the variables of the study. Factor I has its highest loading on the subscales of part 2, whereas factor II has its highest loadings on sections A and B of part 1 and also on religious background. Very roughly, factor I may be defined as "Verbalized Acceptance of the Child," factor II as "Formalized Religious Involvement."

Perhaps the most interesting aspect of the factor analysis is the finding that two of the subscales of part 2 of the questionnaire, Overprotection and Discipline, have appreciable loadings on both factor I and II. To a lesser degree, the same holds true for the Acceptance of Cause subscale. Overprotection has a loading of .43 on factor I, .32 on factor II. Discipline has a loading of .52 on factor I, .40 on factor II. Acceptance of Cause has a loading of .67 on factor I, .25 on factor II.

These results reflect further on the relation between parts 1 and 2 of the questionnaire, between, that is, intensity of formalized religious observance

TABLE 5

INTERCORRELATIONS OF THE VARIABLES EMPLOYED IN THE STUDY

Variables	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
1. Mother's Self-Fulfillment		.42*	.37*	.28*	.41*	.02	.04	.02	.04	−.04	−.14	−.19	−.08	−.01	.11	−.13	.11
2. Discipline			.26*	.33*	.44*	.21	.42*	.11	−.12	.01	−.11	−.24*	−.08	−.07	.12	.13	.06
3. Overprotection				.23*	.39*	.24*	.22	.21	−.08	−.20	−.12	−.02	−.10	−.08	−.03	.08	.05
4. Acceptance of Diagnosis					.50*	−.15	.02	.09	−.20	.16	−.17	−.04	−.01	−.27*	−.20	−.11	.09
5. Acceptance of Cause						.12	.20	.13	−.06	−.12	−.04	−.10	.12	−.07	.12	.13	.22
6. Mother's Religious Observance							.48*	.42*	−.05	−.60*	−.21	−.26*	.16	.17	.27*	.11	.15
7. Father's Religious Observance								.27*	−.14	−.21	−.06	−.13	−.05	−.09	.10	.20	.17
8. Catholic											.02	−.09	.20	.08	−.02	−.11	.15
9. Protestant											−.16	.01	−.11	.01	.14	.21	−.07
10. Jewish											.21	.14	−.13	−.15	−.19	−.15	−.12
11. Mother's Age												.28*	.01	−.04	−.03	−.09	−.03
12. Child's Age													.09	−.02	−.16	−.04	−.20
13. Number of Siblings														.05	.10	.13	.27*
14. Sex of Child															.13	.13	−.09
15. IQ of Child																.34*	−.10
16. Child's Adjustment (judged by pediatrician)																	.21
17. Mother's Acceptance (judged by social worker)																	

* Correlations as great as or greater than ±.23 are significantly different from zero at the .05 level. Correlations as great as or greater than ±.30 are significantly different from zero at the .01 level.

and attitudes of parents of retarded children. Mothers who scored "high" in religious practice tended also to score "high" in verbally expressed attitudes reflecting acceptance, especially when factors of overprotection or discipline were involved.

TABLE 6

FACTOR ANALYSIS OF VARIABLES EMPLOYED IN STUDY

Variables	FACTOR LOADINGS *I*	*II*
Part 2 (Subscales):		
Mother's Self-Fulfillment	.58	.13
Discipline	.52	.40
Overprotection	.43	.32
Acceptance of Diagnosis	.72	—.13
Acceptance of Cause	.67	.25
Part 1 (Sections A and B):		
Mother's Religious Observance	—.11	.75
Father's Religious Observance	.11	.62
Other Variables:		
Catholic	.07	.46
Protestant	—.16	—.13
Jewish	.15	—.56
Mother's Age	—.11	—.20
Child's Age	—.12	—.32
Number of Siblings	.01	.14
Sex of Child	—.25	.15
IQ of Child	—.11	.36
Child's Adjustment	—.10	.30
Mother's Acceptance ($N = 35$)	.15	.23
Mother's Acceptance ($N = 37$)	.07	.06

The following variables, in addition to those already described, had loadings greater than .25 on factor II: IQ of the child (.36), child's adjustment (.30), and child's age (—.31). There is no ready explanation for these findings. One additional variable had a loading of .25 on factor I: sex of the child (—.25). A slight tendency was found for mothers of female children to give responses on part 2 reflecting greater acceptance, consistent with Farber's (5) finding that the marital integration of parents was less affected by the presence of a retarded girl than of a retarded boy in the home.

DISCUSSION

A low but positive correlation was established between measures of religious background and maternal acceptance. The strength of relationship varied with the type of item on the attitude scale which was the measure of acceptance. Items on the attitude scale reflecting dispositions toward discipline of and overdependence in the child correlated more strongly with religious background than other types of items. That is, mothers rating themselves more intense in their religious practices tended to express attitudes judged more acceptant, especially when these attitudes involved habits of discipline or overdependence.

Catholic mothers reported themselves and their husbands as more intense in religious practices than Protestant or Jewish mothers. On the attitude scale as a whole, Catholic mothers responded in a manner judged somewhat more acceptant of their retarded children. This finding gravitates in support of previous findings (5, 15, 27) that Catholic mothers are more acceptant than non-Catholic mothers and contrasts with Boles' (2) finding that Catholic mothers are more guilty, more anxious, and more socially withdrawn. The present finding suggests that maternal acceptance as reflected in verbalized attitudes, as well as in marital integration (5) and institutionalization (15), is influenced *in the same direction* by religious background.

The direction of influence was hypothesized as a result of differences in religious doctrine with respect to the problem of parental guilt. Catholic doctrine is explicit that parents should not feel guilt for bearing a retarded child; indeed, they are asked to accept the child as a special gift of God. Protestant and Jewish doctrine is neither so explicit nor absolving. Emphasis in the Jewish group particularly on intelligence as a special value constitutes an additional handicap in acceptance. Since it has often been noted in the clinical literature that guilt impedes acceptance of the handicapped child, it was believed that differences in acceptance might be detected as a result of the special Catholic injunction that parents not hold themselves personally at fault.

The clinical literature also states that there may be wide discrepancy between acceptance judged in different ways, as occurred for example in the failure to find correlation between measures of clinically-judged acceptance and verbalized acceptance on the attitude scale. The attitude scale, although found to be a fairly reliable measure when evaluated by split-half reliability coefficients of subscales, was not at all predictive when taken in relation to clinical judgment of acceptance as a validating criterion. It is known that attitudes and other behavior of parents toward their handicapped children may vary widely from one area to the next, and at times even be quite self-contradictory. Boles (2) found evidence for this in mothers of cerebral palsied children who proved unexpectedly realistic in appraising their children's mental ability, but ". . . . when asked to estimate their children's future achievements, more basic unrealistic attitudes were revealed" (p. 211). Yet another kind of inconsistency was noted by Zuk (26) as a tendency of mothers to give information about their retarded children which portrayed the children in a better light developmentally than they were portrayed by their teachers. In so doing, the mothers were believed to be satisfying a frustrated need for their children to be more normal in intelligence than they actually were.

The recognition of handicap in their children provokes quite contradictory thoughts and feelings in parents, as Kanner has noted (7). He cited several cases of parents who could not forgive their children for failing to live up to expectations. These parents were intensely frustrated realizing that their children would not fulfill ambitions. How intense this frustration is may be envisioned in Bice's (1) report of two cases in which parents approached him to ask if they might discuss in group psychotherapy how they had once planned to kill their handicapped children. Clearly, certain thoughts and feelings of the parents can be outrightly hostile to the child. But the social setting in which parents live, the culture, frowns on this form of outright hostility, especially in the case of handicapped children. The hostility must, therefore, be kept under lock and key or be expressed in a more socially acceptable manner. It may even be directed against the child if suitably disguised, such as by applying pressure to perform beyond capacity. The psychoanalyst Rickman summarizes succinctly the psychological problem of parents of handicapped children when he asks consideration of the powerful conflict set up by ". . . man's seemingly eternal problem of love and hate for the same person—with ambivalence, which to many is an intolerable burden" (13, p. 125).

The intense conflict, frustration, and ambivalence of parents of retarded children are believed responsible for the difficulty in constructing *predictive* measures of acceptance of the child. This is not to say, however, that such measures do not provide a useful service in revealing clues as to the over-all condition of acceptance. The attempt to treat acceptance as an experimental variable clearly illustrates its many-faceted character. This variable, apparently so unitary and compact in clinical contexts, breaks up into parts and subparts when subjected to methodological treatment. Perhaps the greatest value of research on this variable is to demonstrate its multidimensional character, describe its various levels, and attempt to establish correlates among levels. Methodologically, treatment of acceptance in either molar or molecular interactions, as contrasted recently by Schaefer (17) in a re-evaluation of some maternal behavior concepts, would seem applicable to the general problem envisioned here. It is hoped that the present study has made some contribution in this direction.

SUMMARY

Maternal acceptance of retarded children was explored through means of a questionnaire which tapped attitudes toward the children and religious background. A low but positive correlation was found between measures of maternal acceptance and religious background. Mothers who rated themselves more intense in religious practices tended to verbalize attitudes judged more acceptant of their retarded children. Catholic mothers rated themselves and their husbands more intense in religious practices than non-Catholic mothers and also verbalized attitudes judged more acceptant.

Religious background correlated more positively with maternal acceptance when judgment of acceptance was based on attitude items involving dispositions toward discipline and overdependence.

The interpretation of the findings was believed to gravitate in similar direction to that of other findings which have recently been reported: namely, that Catholic mothers tend to be more acceptant of their retarded children than non-Catholics. Catholic doctrine is more explicit than non-Catholic in that parents consider themselves absolved of personal guilt in the event they bear a retarded or otherwise handicapped child. Since guilt has been assigned a role of major importance in impeding parental acceptance, it was believed that Catholics would be less blocked in acceptance than non-Catholics as a result of the dictum regarding guilt.

The presence of conflict, frustration, and ambivalence in mothers was believed to create special problems in validating a measure of acceptance from the attitude scale, although the measure was found to be reasonably reliable. These psychodynamics will probably continue to make it difficult to devise measures of acceptance based on expression of attitudes which will have much predictive capability. On the other hand, such measures can provide useful clues as to the condition of acceptance when used in comparison with other types of measures and observations.

REFERENCES

1. BICE, H. V. Group counseling with parents of the cerebral palsied. In *Psychological problems in cerebral palsy: a symposium*. Natl Soc. Crippled Child. & Adults, 1952.
2. BOLES, G. Personality factors in mothers of cerebral palsied children. *Genet. Psychol. Monogr.*, 1959, 59, 159-218.
3. BREITENBECK, G. *For parents of retarded children*. Liguori, Mo.: Liguorian Pamphlets, 1958.
4. EATON, J. W., & WEIL, R. J. *Culture and mental disorders: a comparative study of the Hutterites and other populations*. Free Press, 1955.
5. FARBER, B. Effects of a severely retarded child on family integration. *Monogr. Soc. Res. Child Develpm.*, 1959, 24, No. 2.
6. HASTINGS, D. Some psychiatric problems of mental deficiency. *Amer. J. ment. Def.*, 1948, 52, 260-262.
7. KANNER, L. The emotional quandaries of exceptional children. In *Helping parents understand the exceptional child*. Langhorne, Pa.: Woods School, 1952. Pp. 21-28.
8. MILLER, R. L. Religious dimensions of maternal acceptance of the retarded child. Unpublished master's thesis, Princeton Univer., 1959.
9. MOORE, C. A study of the parental attitudes towards cerebral palsied children. Unpublished master's thesis, Univer. of Michigan, 1953.

10. Perlstein, M. A. *The present-day status of the cerebral palsy problem.* Assn. Aid Crippled Child., New York City, 1947.

11. Ray, D. A study to develop a guide of education for parents of cerebral palsied children. Unpublished master's thesis, Univer. of Iowa, 1951.

12. Rheingold, H. Interpreting mental deficiency to parents. *J. consult. Psychol.,* 1945, 9, 142-148.

13. Rickman, J. *Selected contributions to psychoanalysis.* Basic Books, 1957.

14. Roe, H. The psychological effects of having a cerebral palsied child in the family. Unpublished doctoral dissertation, Teach. Coll., Columbia Univer., 1950.

15. Saenger, G. *The adjustment of severely retarded adults in the community.* Report to New York State Interdepartmental Health Resources Board, 1957.

16. Sarason, S. *Psychological problems in mental deficiency.* (2nd. Ed.) Harper, 1953.

17. Schaefer, E. S. A circumplex model for maternal behavior. *J. abnorm. soc. Psychol.,* 1959, 59, 226-235.

18. Schaefer, E. S., & Bell, R. Q. Development of a parental attitude research instrument. *Child Develpm.,* 1958, 29, 339-361.

19. Scheimo, S. L. Problems in helping parents of mental defective and handicapped children. *Amer. J. ment. Def.,* 1951, 56, 42-47.

20. Sears, R. R., Maccoby, E. E., & Levin, H. *Patterns of child rearing.* Row, Peterson, 1957.

21. Stone, M. M. Parental attitudes to retardation. *Amer. J. ment. Def.,* 1948, 53, 363-372.

22. Thurstone, L. L. *Multiple-factor analysis.* Univer. of Chicago Press, 1947.

23. Walker, H., & Lev, J. *Statistical inference.* Holt, 1953.

24. Waterman, J. H. Psychogenic factors in parental acceptance of feebleminded children. *Dis. nerv. System,* 1948, 9, 184-187.

25. Wishik, S. M. *How to help your handicapped child.* Publ. Affairs Committee, 1955, No. 219.

26. Zuk, G. H. Autistic distortions in parents of retarded children. *J. consult. Psychol.,* 1959, 23, 171-176.

27. Zuk, G. H. The religious factor and role of guilt in parental acceptance of the retarded child. *Amer. J. ment. Def.,* 1959, 63, 139-147.

RECOGNITION OF PERCEPTUALLY AMBIGUOUS STIMULI IN GRADE SCHOOL CHILDREN

JURIS G. DRAGUNS*
Rochester State Hospital

and GIUSEPPE MULTARI[1]
University of Rochester

Since Frenkel-Brunswik (4) introduced the concept of intolerance of ambiguity as a combined personality, perceptual, and cognitive variable, individual differences in readiness or reluctance to invest meaning into poorly structured stimuli have been employed as one of its operational indices (5, 9, 10, 13, 14). However, there are still no systematic data available on the manner in which these presumably constant and personality related response tendencies develop in childhood and adolescence. While a study by Korchin and Basowitz (6) has explored the effects of aging on perception and judgment of ambiguous visual stimuli, no such investigations have been conducted with children at various age levels. Several reports concerned with intolerance of ambiguity made use of preadolescent and adolescent subjects (9, 10, 14), but in none of these were there any age comparisons made.

More directly relevant to the problem at hand is the investigation by Dalla Volta (2) whose interest was focused on the ability of children between the ages of 4 and 12 to describe and reproduce pictorially various kinds of "naturally indistinct" stimuli, e.g., clouds. Dalla Volta's principal finding points to small age differences and to an unquestioning and presumably nonanxious attitude of his subjects in dealing with these expectedly and naturally vague parts of their environment. His report raises two questions appropriate to this investigation: Does this kind of behavior hold when a child is confronted with an artificially induced, unexpectedly ambiguous stimulus? Would age differences be more prominent if the subjects were

---

* Department of Psychology, Rochester State Hospital, 1600 South Avenue, Rochester 20, New York.

[1] Now at the University of Ottawa.

Child Develpm., 1961, 32, 541-550.

faced with "artificial indistinctness" induced by one of the many techniques of stimulus impoverishment?

Two other research reports throw some light on these issues. L'Abate (8) found, upon comparing the responses of children from kindergarten through 5th grade to a multiple choice picture story test and to some simple animal pictures, an increase in the number of "uncertainty expressions" and qualifiers as a function of age. Piaget and Stettler - von Albertini (12) reported that incomplete or overlapping geometrical figures tend to give rise among 4- to 6-year-olds to perception of "empirical configurations" such as houses, people, and utensils while 7- to 8-year-olds more veridically described such stimuli in terms of various geometrical and other abstract shapes. We interpret this finding as possibly indicating a greater tendency toward early —and maybe premature—meaningful structuring of visual stimuli at an early age. The results of both L'Abate and of the Swiss investigators would suggest that early age levels are characterized by quick and conflict-free recognition of ambiguous stimuli, unimpeded apparently by critical appraisal of the scant bases for such judgments. To explore further the relevance of age factors in children to the attribution of meaning to ambiguous stimuli is the object of the current study.

SUBJECTS

Thirty *S*s each, consisting of an equal number of boys and girls, were selected randomly from various Catholic parochial schools in Rochester, New York,[2] at the following grade levels: 1st, 3rd, 5th, and 7th. These grades were chosen so as to reduce overlap in chronological age and still have a reasonably continuous picture of the behavior of children of various ages within the grade school range when confronted with our stimuli.

To explore the intellectual composition of our groups, an appropriate form of the Kuhlmann-Anderson intelligence test was group administered to a given grade level. The obtained means and standard deviations of IQs, summarized in Table 1 along with data on the *S*s' age, show that the grade and sex groups are largely comparable in intelligence as no differences among them approach statistical significance.

MATERIALS AND PROCEDURE

The task selected to assess our *S*s' meaningful structuring of ambiguous stimuli follows largely the "decision location" format employed in the study of recognition response level by Binder (1), Messick and Hills (10), Smock (13, 14), and others. In all of these, a series is presented starting with a

---

[2] Sincere appreciation is expressed to Rev. William M. Roche, Diocesan Superintendent of Schools, for permitting us to examine our subjects and to Sisters M. Beatrice, SSND, Gertrude Marie, SSJ, and Helen Therese, SSJ, for assistance in making classes available for our experiment.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF AGES AND INTELLIGENCE QUOTIENTS IN GRADE AND SEX GROUPS

	Age		Intelligence	
	M	*SD*	*M*	*SD*
Grade I				
Boys	6.9	0.61	105.2	7.5
Girls	6.7	0.36	109.4	4.1
Total	6.8	0.56	107.3	6.5
Grade III				
Boys	8.2	1.27	116.7	15.2
Girls	8.0	0.65	114.3	13.8
Total	8.1	0.78	115.5	14.8
Grade V				
Boys	9.8	0.64	103.1	8.3
Girls	9.8	0.65	99.1	9.1
Total	9.8	0.64	101.1	8.9
Grade VII				
Boys	12.6	0.74	100.7	7.7
Girls	12.6	0.42	106.8	9.1
Total	12.6	0.64	103.8	8.8

stimulus which, perceptually or cognitively, conveys minimal information to the subject. The series then proceeds, with appropriate graduations, to a presentation which permits unhampered recognition of the object or contains all of the information for the solution of the problem at hand. To construct such a series, various methods of obscuring, camouflaging, or incompletely presenting the stimulus have been employed. In our case, we followed a procedure applied by Voigt (15) in a different context. For our purposes, the following nine drawings were made: a sail boat, a bridge, a dog, a bowl of fruit, a house, a telephone, a group of trees, a violin, and a wheelbarrow, and then photographed with the objective progressively moved out of focus.[3] In this manner, at least 12 pictures of each object were prepared. These were presented to 10 adult judges to see if their ranking of the pictures from unclear to clear would correspond to the degree of actual photographic blurring. On the whole this correspondence was high. The few photographs on which the judges' rankings were inconsistent were eliminated, but no other changes were made. Thus for experimental use, nine series of 12 pictures each were retained which ranged, in most cases, from the representation of a formless, blurred mass to a virtually unequivocally perceptible picture of a common object.

---

[3] The help of Miss Virginia Jeffrey Smith in providing the drawings for our experiment is gratefully acknowledged. We also thank Mr. Victor Atyas for his help in photographing our stimuli.

Each series of pictures was arranged in a booklet. These booklets were presented to the subjects in a random order to eliminate possible learning, habituation, or fatigue effects. The following instructions were given:

> I have several booklets of pictures to show you. Some of the pictures in these booklets are kind of blurry. As we go through the booklets, I would like you to tell me when you feel you have an idea what the picture is. So keep in mind that it's okay to stop me at *any* point when you feel you have an idea what the picture is.

Each picture was shown to the subject for 10 seconds at the end of which time *E* turned the page, thus revealing the next picture. Presentation time was restricted to prevent perseverations and large variations in reaction time (which are not to be explored in this study), but at the same time to allow a fairly complete and unhurried examination of the card. With this procedure, all reactions occurred within the 10-second time limit, and wide variations in reaction time were not observed. The entire administration was completed within 10 to 20 minutes in most cases.

Presentation of each series yielded us one response with the following kinds of information: (a) the number of the picture, from 1 (at the vague end of the series) to 12, on which the *S* ventured to report seeing some sort of a concrete object; (b) whether or not the object seen was the one presented clearly on the final picture series. The former was used to compute the recognition level score; the latter was included in the total tally of correct or error responses, respectively. The *S*'s recognition level score

TABLE 2

SPLIT-HALF CORRELATION COEFFICIENTS FOR RECOGNITION LEVEL AND ERROR SCORES

	Recognition	*Errors*
1st grade	.78	.27
3rd grade	.65	.68
5th grade	.39	.21
7th grade	.85	.37
Total	.73	

NOTE.—Pearson *r*s were computed for recognition level and Spearman *rho*s for errors.

was so derived for each series, and an individual's total score was found by averaging his nine obtained scores. The number of errors, only incidental to the main purpose of this investigation, was likewise computed for each individual.

To obtain an estimate of the reliability of *Ss'* recognition level and error scores, split-half correlation coefficients were computed by comparing their performance on the first five and last four series. As will be seen from the examination of Table 2, reliability is, on the whole, moderately high for the recognition responses of various grade groups as well as for the total population. Error scores, however, are much less consistent, it would seem.

RESULTS

Recognition level scores were investigated by means of a two-way analysis of variance for sexes (2) by grades (4). Table 3 summarizes the results of this analysis and reveals highly significant effects due to grade level ($p < .01$) while effects of both sex and sex-grade interaction are negligible. Comparison of mean grade level scores by means of t tests yields only one difference at the .05 level of confidence, that between the 1st and the 5th grade. Differences between 1st and 7th and between 3rd and 5th grades are of borderline significance ($p < .10$) in both cases.

Post hoc examination of recognition level means and standard deviations shown in Table 4 suggests a rather pronounced contrast between the two lower and the two higher grades, an impression which is confirmed by statistical analysis ($p < .05$).

Since distribution of error scores deviated markedly from normality, a distribution-free chi square analysis of variance, as described by Wilson (17), was applied to these data. This procedure, again, revealed significant grade effects at the .01 level, but no other reportable findings. Mann Whitney *U* tests performed to compare differences among various grade levels resulted in the following significant findings at the .05 level of confidence:

TABLE 3

ANALYSIS OF VARIANCE FOR RECOGNITION LEVEL IN SEX AND GRADE GROUPS

	df	*MS*	*F*	*p*
Grade	3	96.24	18.43	.01
Sex	1	6.85	1.28	*ns*
Grade × Sex	3	6.48	1.22	*ns*
Within Cells	112	5.27		

1st grade vs. 5th grade, 1st vs. 7th, and 3rd vs. 5th grade. Comparison of means and standard deviations of error scores as shown in Table 4, though possibly misleading in view of their skewed nature, points to a well-nigh continuous decline in the number of errors from our youngest to our oldest group while a tendency to a later object recognition increases with age. On this basis and on logical grounds, one would expect a sizeable negative correlation between errors and recognition level; the earlier an attempt at object naming, the more of a chance there is for making a mistake in view of the incomplete information available. This expectation is, by and large, borne out by the data in Table 5. More noteworthy perhaps are the results of correlating these two mutually associated measures with intelligence. While recognition scores consistently show a negligible correlation with the IQ level, there is, in some grades at least, a significant negative correlation of errors with intelligence. This finding parallels that of Smock (14) who reported no relation between intelligence and the *S*'s first response to an analogous task while a significant correlation was obtained between intelligence and the first correct response.

TABLE 4

MEANS AND STANDARD DEVIATIONS OF RECOGNITION LEVEL AND ERROR SCORES IN GRADE AND SEX GROUPS

	RECOGNITION		ERRORS	
	M	*SD*	*M*	*SD*
Grade I				
Boys	3.0	1.1	8.4	1.1
Girls	4.8	2.6	7.2	3.1
Total	3.9	2.1	7.8	1.9
Grade III				
Boys	4.6	2.4	5.9	2.7
Girls	4.4	2.9	5.3	3.0
Total	4.5	2.7	5.6	2.9
Grade V				
Boys	7.6	0.9	4.6	1.7
Girls	8.0	2.9	4.5	2.1
Total	7.8	1.9	4.6	1.9
Grade VII				
Boys	6.7	2.4	4.7	2.6
Girls	7.2	1.9	5.1	1.9
Total	7.0	2.2	4.9	2.1

TABLE 5

CORRELATION COEFFICIENTS AMONG RECOGNITION LEVEL, ERRORS, AND INTELLIGENCE QUOTIENTS

	Recognition Level *with* Errors	Recognition Level *with* Intelligence	Errors *with* Intelligence
1st grade	—.45**	.18	—.44*
3rd grade	—.74**	.12	—.42*
5th grade	—.51**	.06	—.32
7th grade	—.69**	.12	—.17
Total		.09	

NOTE.—Spearman *rhos* were computed, instead of Pearson *rs*, for all correlations with error scores.

* Significant beyond .05 level.

** Significant beyond .01 level.

DISCUSSION

Within a limited age span, our results seem to point to an increased cautiousness in structuring vaguely seen stimuli as a function of either higher age or greater educational attainment and to a concomitant decrease in the number of errors. The absence of sex differences may be worth noting in view of L'Abate's reporting such differences in his previously cited project (8). However, since the above author was investigating variations in verbal behavior while we were concerned with perceptual recognition, we can merely report this possible discrepancy without speculating at this time on the generality or specificity of the findings which gave rise to it.

Our positive results should be interpreted with caution. We do not yet know whether a continuous trend toward more conservative and more accurate judgments would hold true for the entire span from preschool age through adulthood or whether the small and statistically insignificant reversal to lower recognition level scores in the 7th grade may presage the articulation of such a tendency in adolescence. Furthermore, the results obtained could be attributed to either age or educational level or a combination of both. Perhaps, they may be regarded as a function of all relevant effects, educational and otherwise, which have impinged upon an individual up to the time of our experiment.

With the above limitations in mind, our data could be explained from two different, but overlapping, points of view. Within the developmental framework, our younger *Ss*' tendency toward early and often incorrect identification of picture contents might be interpreted, in Werner's (16) terms, as an instance of syncretism, i.e., a lack of segregation between internally and externally determined experiences. If we adopt the microgenetic approach (3, 7, 15) and assume a continuous perceptual progression from the initial presentation to final structuring, the process of recognition may be viewed as a gradual balancing of subjective determinants and objective clues. The *S*'s choice is then between extending the exploration of the picture —i.e., to use Piaget's terms, (11) "accommodating" himself to its emerging properties—or cutting this search short by "assimilating" the percept to the objects of his previous experience. The point on a continuum from indistinctness to clarity at which such decisions are made constitutes our basic measure and, on the basis of our findings, seems to differ with age and/or educational level.[4] Our findings, however, in no way refute the possibility—and the relatively wide range of differences within our age groups indirectly tends to support it—that personality factors may contribute, along with developmental ones, to determining our *S*'s scores. Our interpretation then need not dispense with the assumption either explicitly or implicitly made in many studies of intolerance of ambiguity, viz., that perceptual—and other—kinds of ambiguity are, to varying degrees, productive of anxiety (4, 14). Italian research (2, 18) and our own episodic impressions, however, suggest that children's reactions to poorly structured parts of the environment are free of at least overt discomfort.

In this, there seems to be a certain contrast between our data and those obtained by Korchin and Basowitz (6) on the judgment of ambiguous stimuli in the course of aging. While their experimental materials and procedures are not fully comparable to ours, they noted appreciable differences in responses to situational ambiguity—inherent in the incomplete structuring of the purpose and task of the experiment—in contrast to minimal differences in the readiness to judge, and the correctness of judging, actual ambiguous experimental materials. Integrating Dalla Volta's (2) and Zecca's (18) results with ours, we see a suggestion of a reverse tendency in children; while the ambiguities of the experimental task are accepted unreflectively and nonanxiously, stimulus recognition as such gives rise to perceptible age differences. Future research may investigate the hypothesis that development in childhood is characterized by changes in responding to stimulus ambiguity, but not to situational ambiguity.

Finally, it may be tempting if premature to try integrating our results with the personality oriented literature on intolerance of ambiguity. Thus, tendencies by anxious individuals toward an accelerated appearance of attempts at labeling an object and imbuing it with meaning (14) may conceivably be conceptualized as regression to modes of responding characteristic of earlier stages of development. Difficulties, however, arise in pursuing this line of reasoning as the tendency of avoiding committal to material which is open to several interpretations has also been found among anxious *Ss* (5, 13, 14). Perhaps, the latter kind of response tendency may represent direct effects of anxiety while the former may be akin to a defense against it. Again, future investigations only can clarify this, at the present time, quite speculative point.

---

[4] A recent report by J. Drösler and W. F. Kuhn, which came to our attention after preparing this article, points out, on the basis of perceptual and psychophysical experiments, a greater tendency toward symmetry, structuring, and closure in children as compared with adults and consequently parallels to a certain extent our findings. *See* Drösler, J., & Kuhn, W. F., Ein experimenteller Vergleich der visuellen Wahrnehmung von Kindern, Schizophrenen und Alkoholikern mit der tachistoskopischen Wahrnehmung normaler Erwachsener. *XVI International congress of psychology. Individual papers. Summaries.* Bonn: German Society of Psychology, 1960, I, 1 - I, 2.) Less directly related but still relevant are the results of F. Winnefeld (Gestaltauffassung und Umgestaltung in genetischer Sicht, *Z. exp. angew. Psychol.*, 1959, 6, 589-602) who reports increases in the ability to find hidden figures and to reverse figure and ground as a function of increasing age. This appears to be in line with a more object determined response to ambiguous stimuli observed by us.

SUMMARY

Children of both sexes in four different grades were compared as to their readiness to attribute meaning to pictures differing in the amount of photographic blurring. Consistent if small differences among grades suggested a tendency toward a greater amount of cue accumulation before making such judgments as a function of age and/or grade level. The number of errors on this task likewise declined at the higher grade levels. Relevance of these findings within a developmental framework was discussed, and possible parallels as well as differences were noted upon the comparison of our data with the performance of aged individuals and of anxious subjects on similar tasks.

REFERENCES

1. BINDER, A. Personality variables and recognition response level. *J. abnorm. soc. Psychol.,* 1958, 57, 136-142.
2. DALLA VOLTA, A. Rappresentazione dell'indistinto a sviluppo del bambino. *Minerva Pediat.,* 1951, 3, 44-79.
3. FLAVELL, J. H., & DRAGUNS, J. A microgenetic approach to perception and thought. *Psychol. Bull.,* 1957, 54, 197-217.
4. FRENKEL-BRUNSWIK, E. Intolerance of ambiguity as an emotional and perceptual variable. *J. Pers.,* 1948, 18, 104-141.
5. HAMILTON, V. Perceptual and personality dynamics in reactions to ambiguity. *Brit. J. Psychol.,* 1957, 48, 200-215.
6. KORCHIN, S. J., & BASOWITZ, H. The judgment of ambiguous stimuli as an index of cognitive functioning in aging. *J. Pers.,* 1956, 25, 81-95.
7. KRAGH, U. *The actual-genetic model of perception-personality.* Lund: C. W. K. Gleerup, 1955.
8. L'ABATE, L. Sanford's uncertainty hypothesis in children. *Etc. Rev. gen. Semant.,* 1957, 14, 210-221.

9. Levitt, E. E. Studies in intolerance of ambiguity: I. The decision location test with grade scholl children. *Child Develpm.*, 1953, 24 263-268.

10. Messick, S., & Hills, J. R. Objective measurement of personality: cautiousness and intolerance of ambiguity. *Educ. psychol. Measmt.*, in press.

11. Piaget, J. *The origins of intelligence in children.* International Universities Press, 1952.

12. Piaget, J., & Stettler - von Albertini, B. Recherches sur le développement de perception XIX. Observation sur la perception des bonnes formes chez l'enfant par actualization des lignes virtuelles. *Arch. Psychol* (Genéve), 1954, 34, 203-242.

13. Smock, C. D. The influence of psychological stress and the "intolerance of ambiguity." *J. abnorm. soc. Psychol.*, 1955, 50, 177-182.

14. Smock, C. D. The relationship between "intolerance of ambiguity," generalization, and speed of perceptual closure. *Child Develpm.*, 1957, 28, 27-36.

15. Voigt, J. Die Aktualgenese in der psychologischen Diagnostik. *Psychol. Beitr.*, 1956, 2, 586-636.

16. Werner, H. *Comparative psychology of mental development.* International Universities Press, 1957.

17. Wilson, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, 53, 96-101.

18. Zecca, G. Il comportamento grafico del bambino nelle prime rappresentazioni degli indistinti naturali. *Arch. Psicol. Neurol. Psichiat.*, 1956, 12, 736-754.

QUANTITY CONCEPTIONS IN JUNIOR AND SENIOR HIGH SCHOOL STUDENTS[1]

David Elkind*
UCLA School of Medicine

In Piaget's (11) experiments dealing with the development of quantity conceptions in children, he found that abstract responses to mass, weight, and volume appeared in a regular sequence that was related to age. Abstract responses to mass appeared (i.e., in 75 per cent of the subjects tested) by ages 7 to 8, abstract responses to weight appeared at ages 9 to 10, and abstract responses to volume by ages 11 to 12. A systematic replication of one of Piaget's experiments (6) confirmed Piaget's ages for mass and weight but failed to confirm his age for volume which was conceived abstractly by only 27 per cent of 11- to 12-year-old American children. The present study seeks to extend the replication to 12- to 18-year groups and to determine the influence of age, sex, and IQ on the attainment of abstract conceptions of quantity in adolescents.

According to Piaget, the criterion of an abstract conception of quantity is the child's judgment of its conservation, its sameness despite perceptual change. For example, in one of Piaget's experiments he showed his subjects

---

* Department of Psychiatry, School of Medicine, University of California Medical Center, Los Angeles 24.

[1] This investigation was supported (in part) by a PHS Small Grant M-3466 from the Department of Health, Education, and Welfare, U. S. Public Health Service; the author was then at Wheaton College.

Child Develpm., 1961, 32, 551-560.

two identical clay balls, one of which was subsequently rolled into a "sausage." Those children who judged the mass, weight, or volume of the "sausage" as conserved (the same) demonstrated an abstract conception of quantity. The same technique and criterion were used in the present study.

METHOD

Subjects

Four hundred and sixty-nine junior and senior high school students attending the Norton High School in Norton, Massachusetts, were tested.[2] The majority of children came from lower-middle class families. The mean age and IQ for boys and girls at each grade level are given in Table 1.

Tests and Procedure

The tests for conservation were given as group[3] tests, and the students wrote their answers in blue books. Materials for all tests were the same: two identical clay balls (1½ inches in diameter) and a small scale. The tests were always given in the following order.

1. *Test for the conservation of mass.* E began by explaining that the two balls were identical in every way, that there were no tricks in the experiment, and that he was simply repeating an experiment originally performed with elementary school pupils. Several students were asked to verify by means of the scale that the two balls weighed the same. E asked any student who still had doubts to voice them. Questions raised permitted E to explain further that the balls were made of the same kind of clay and that "same amount" meant to the degree of accuracy permitted by the scale and did not mean absolute equivalence.

Once the students agreed that the balls were the same, E asked: (a) "Do the balls both contain the same amount of clay?" (identity question); (b) "Suppose I made one of the balls into a sausage, would the two pieces of clay still contain the same amount of clay?" (prediction question). E then actually rolled one of the balls into a sausage and asked: (c) "Do they both contain the same amount of clay now?" (judgment question); (d) "Explain your answers" (explanation question).

2. *Test for the conservation of weight.* E rolled the "sausage" back into a ball and then repeated the whole procedure described above save for the use of "weight" instead of "amount."

3. *Test for the conservation of volume.* E again rolled the "sausage" back into a ball and repeated the procedures for mass and weight, except that the terms "volume" and "same room or space" were substituted for the previously used quantity concepts.

---

[2] The writer is indebted to Mr. Henri A. Yelle, Principal of the Norton High School, and to the teachers and students whose friendly cooperation made the study possible.

[3] In Piaget's (12) experiments and in the first replication (6) with elementary school children, the tests were administered individually.

Controls

Objections were raised by Baldwin (2) against Dennis' (3) finding of animism in college students. The brunt of these objections was that verbal misunderstandings and verbal ambiguities made interpretation hazardous. In the present study a number of controls were introduced to minimize the verbal misunderstanding and ambiguity: (a) the preliminary procedures of weighing the balls and of answering questions regarding the experiment tended to reduce misunderstandings; (b) the four types of question asked for each concept—identity, prediction, judgment, and explanation—gave some measure of the reliability of responses; (c) finally, as a special check on the conservation of volume test (which proved most difficult for children in the original and replication study), students were asked what would happen if the ball and the sausage were put into identical glasses filled equally high with water.

Scoring

A test was considered passed only if the subject correctly identified, predicted, judged, and explained the conservation of a quantity.

RESULTS

Controls

The control tests indicated that students who failed the conservation tests did so because of inadequate conceptions and not because of verbal misunderstanding.

All of the students who failed the volume test agreed (answered "Yes" to the identity question) that initially the balls were the same in volume. But they predicted, judged, and explained that the volume of the clay changed with the change in its shape. This was true for the same students who on the previous pages had written that mass and weight were conserved because either: (a) nothing was added or taken away; or (b) changing the shape did not change the amount; or (c) what the sausage lost in width it gained in length and therefore was the same as the ball.[4] Thus, the students clearly understood the conservation questions with regard to mass and weight, but failed to generalize their judgments, predictions, and explanations to the conservation of volume which they treated as an entirely different problem. Kay (16-5) wrote, for example, that mass and weight were conserved because nothing was added or taken away but to the volume question she wrote, "The molecules may be more compressed in one object (the ball) than in the other (the sausage); although it (the sausage) has the same number of molecules and the same weight, its volume is not the same."

The question about the water levels, introduced as a control, revealed a fact significant in itself. Over two thirds (70 per cent) of the students said that the water levels would rise because of the weight (often described as impact or pressure) of the submerged objects. Gla (15-4) "The water with the ball will rise more because the impact (*sic*) of the ball is greater because it is more condensed and it will rise more."

---

[4] These are the same three explanations given by the elementary school children.

Such examples could be multiplied, but even these are sufficient to show that students who failed to judge volume conservation did so, not because of verbal misunderstanding, but rather because they failed, in Piaget's terms, to dissociate their subjective sensorimotor conceptions (impact, compression) of weight and volume from their objective (molecules), logico-mathematical conceptions.

Percentage of Subjects Judging Conservation

Of the 469 junior and senior high school students tested, 87 per cent had abstract conceptions of mass and weight, but only 47 per cent had an abstract conception of volume. A test of the difference between the three percentages gave an χ^2 of 254.2 which is significant beyond the .001 level. Table 1 shows that, according to the usual 75 per cent criterion for placing a test at a given year level, the volume test is too difficult for all but the

TABLE 1

PERCENTAGE AND MEAN IQ OF STUDENTS, CLASSED BY SEX AND AGE, HAVING ABSTRACT CONCEPTIONS OF QUANTITY

				QUANTITY CONCEPTION		
Sex	*N*	*Mean Age*	*Mean IQ*	*Mass*	*Weight*	*Volume*
M	56	12.6	94.6	79	71	38
F	66	12.6	99.2	79	86	26
M	46	13.6	89.6	78	74	43
F	48	13.6	100.9	90	92	29
M	28	14.6	99.6	93	96	68
F	45	14.6	107.9	93	93	40
M	31	15.5	95.3	97	90	58
F	41	15.7	101.9	98	93	39
M	29	16.4	100.0	83	90	72
F	43	16.6	107.9	86	86	58
M	14	17.7	109.9	100	100	79
F	12	17.7	108.3	100	95	68
			TOTALS			
M	204	15.1	96.5	85	83	54
F	265	15.1	104.3	89	90	40
M & F	469	15.1	100.4	87	87	47

oldest age group. This result, like that of the first replication, fails to confirm Piaget's age (years 11 to 12) for the abstract conception of volume.

Variables Influencing Volume Conservation

Age. As shown in Table 1, there was a regular increase with age in the percentage of students who had abstract volume conception.

Sex. Table 1 shows that for each age level the percentage of boys was consistently higher than the percentage of girls who had an abstract conception of volume. For the two groups as a whole the difference gave an χ^2 of 7.9 which is significant at the .01 level.

IQ. The point biserial coefficient for the correlation of IQ (Kuhlmann-Anderson) with passing the volume test was .31, significant beyond the .01 level. This finding agrees with the results of other studies (4, 5, 6) which have also shown a low but positive correlation of IQ with measures of conceptual development.

Discussion

The results of the present study raise two interrelated questions: (a) Why do significantly more students attain abstract conceptions of mass and weight than attain the abstract conception of volume? (b) In what respect are age, sex, and IQ related to volume conservation? Both questions can be answered within the framework of Piaget's genetic psychology.

In the course of more than 40 years of continuous research on mental development Piaget has evolved a genetic theory of intelligence, adaptive thinking and action. He derives, in a necessary sequence of age-related stages, the abstract intelligence of adults from the sensorimotor coordinations of infants. It is a nature theory in the sense that the stages—sensorimotor (0 to 2 years); preoperational (2 to 7 years); concrete operational (7 to 11 years); formal operational (11 to 15 years)—are assumed to reflect maturational changes in the forms of thinking. It is also a nurture theory in the sense that the gradual attainment of particular conceptions is assumed to reflect the influence of physical and social environment on the contents of thought.

Unlike reinforcement learning theorists, Piaget holds that there is no real separation between drives and cognitive structures: "À tous les niveaux, les intérêts et les besoins qui son intervenus dans ces apprentissages étaient eux mêmes solidaires des structures déjà construites ou en voie de construction" (7, p. 30). For Piaget the appearance of new structures gives rise to new drives to exercise those structures. Such exercise is self-reinforcing through the production of function pleasure. In the course of being exercised, novel stimulation leads to the differentiation of new structures which in turn give rise to new drives in a continuous progression.

The attempt by the organism to use the stimulation of the environment to develop and conserve its structures, Piaget speaks of as assimilation. The modification of structures due to environmental influence, he speaks of as accommodation. Any particular cognitive system attains its final stage of development when assimilation and accommodation are in equilibrium. The two processes are in equilibrium when novel stimulation can be used by the system without either disrupting its activity or necessitating a structural differentiation. According to Piaget, the attainment of the equilibrium stage is the underlying dynamic of all intellectual growth.

Applied to quantity conceptions, Piaget's theory holds that the child is ready[5] for attaining the abstract volume conception at ages of 11 to 12. The child is ready in the maturational sense because the conceptualization of volume requires only concrete operations (internalized actions) which are present in most children by the age of seven (5, 6, 11, 12). The child is also ready in the experiential sense that he has had sufficient object contacts to form abstract conceptions of mass and weight which are the structural prerequisites for the attainment of the volume conception (6, 12). On the basis of his findings Piaget (12) assumed that this operational and structural readiness was sufficient for the attainment of the volume conception by a majority of 11- to 12-year-old children. His theory, on the contrary, suggests that a majority of young people should not, as in the present study they did not, attain an abstract conception of volume.

Piaget's theory proposes two kinds of events which when taken in combination can account for the results of this investigation. According to Piaget, the age at which a young person is ready to attain the volume conception is also the age at which he is developing formal mental operations. Whereas concrete operations (7 to 11 years) are concerned with classifying, relating, and quantifying the immediate environment, formal operations are concerned with constructing systems and theories. Put simply, concrete operations are concerned with immediate reality, formal operations with possibility (13).

Because of the unity of operations and drives, the appearance of formal operations means the appearance of new interests and horizons:[6] "The adolescent is no longer content to live in the interindividual relations offered by his immediate surroundings or to use his intelligence to solve problems of the moment" (13, p. 341). The result of the new interests and motivations arising from formal operations is necessarily an attenuation in the drive for conceptualizing the immediate physical environment. This attenuation of drive for inductive conceptualizing therefore decreases the possibility of children spontaneously discovering the conservation of volume.

Intimately bound up with the appearance of formal operations is a second event which affects the early adolescent's conceptualizing activity. This second event is the adoption, beginning around ages 11 to 12, of adult roles. In Piaget's view, roles are systems of values and ideals which serve to control and direct the individual's behavior. Roles are adopted only in early adolescence because ideals and values are possibilities, not realities, and therefore can only be constructed by means of formal operations.

---

[5] This is not Piaget's term but economically describes the child's state.

[6] This may be one of the reasons for the well established decline of interest in school at about the same age (9, 10, 14). Perhaps, the school curriculum does not take sufficient account of the early adolescent's need for hypothetical propositions.

Prior to the adoption of adult roles, conceptualization was directed by the activity of concrete mental operations and the problems posed in structuring the physical environment. With the appearance of formal operations and adult roles at ages 11 to 12, these come to play the leading part in the direction of conceptualizing activity.

With social roles directing cognitive activity, the adolescents' conceptualizing becomes much more selective than it was during childhood. Unlike the relatively constant physical environment which poses similar problems for all youngsters, the diversity of social roles makes for enormous variety of concept formations among adolescents. For example, the young person who wishes to become a scientist will meet a different set of experiences—and hence conceptual problems—than will another young man whose aim is to be a mechanic. And, in general, adolescents will attain different conceptions according to the kinds of experiences and conceptual problems their particular roles provide. Accordingly, those adolescents who attain the volume conception, despite the attenuation of motivation, would have adopted roles conducive to the formation of quantity conceptions.

These two events—(a) the attenuation of interests and motivations for quantity conceptualizing due to the development of formal operations and (b) the increased selectivity, at the same age, of conceptualizing due to the adoption of diverse social roles—can explain why so many young people who spontaneously discovered the conservation of mass and weight nevertheless failed to discover the conservation of volume. It remains to determine whether this conclusion is consistent with the relations of age, sex, and IQ to volume conservation.

Age. The increase with age in the percentage of students having abstract volume conceptions is misleading because the number of students who dropped out of school also increased with age (Table 1).[7] The groups in the higher grades were therefore a much more select sample than were the groups in the lower grades. What the relation of age and volume conservation shown in Table 1 really means is that those young people who remain in school are more likely than those who leave school to have attained an abstract conception of volume.

This finding is in agreement with the hypothesis that individual differences in social roles leads to selective conceptual learning. Students who stay in school are more likely than those who leave to have adopted professional or academic occupational roles. It seems reasonable to assume that the motivations and interests associated with the adoption of such roles would be more likely than nonacademic or nonprofessional roles to lead individuals into situations requiring the conceptualization of quantity.

---

[7] School records showed that the *N* for the 7th grade was average for the initial size of the groups at the older age levels. The difference between the number of 7th graders and the number of the higher grades indicates the number of dropouts at each grade level.

Sex. The consistently higher percentage of boys than girls having an abstract volume conception can hardly be attributed to innate differences in conceptual ability between the sexes. For one thing, the mean IQ of girls is higher than that of boys at each age level. Secondly, there is no difference between sexes with respect to the attainment of mass and weight conceptions. It seems very unlikely that there is an ability specifically for the attainment of the volume conception. Finally, Hurd (8) has shown that even when girls are far behind boys in their knowledge of physics they are almost completely able to catch up with additional training. For all these reasons differences in conceptualizing ability between sexes seems an improbable explanation for the difference in their performance.

The performance differences between boys and girls are consistent with the hypothesis presented earlier. The roles adopted by the American girl differ considerably from the roles adopted by the boy (1). Girls are supposed to be more interested in social relations than boys and to find mathematics and physics incomprehensible.[8] Boys, on the other hand, are supposedly awkward in social situations and intuitively adept at science and mechanics. Many more boys than girls have experiences in measuring and building things, and most boys and few girls get scientific and mechanical books and toys as gifts. This difference in experiences and opportunity for developing quantity conceptions could account for the fact that more boys than girls attain an abstract conception of volume.

IQ. Individual differences in social roles could also explain the low but positive correlation of IQ with volume conservation. The IQ is neither a necessary nor a sufficient condition for the adoption during adolescence of any particular occupational role. Children with high IQs may never adopt a professional or academic ideal, while youngsters with low IQs may set their hearts on being doctors or lawyers. With the adoption of the appropriate role, even adolescents with MAs of 8 or 9 could attain the volume conception since it can be formed with the mental operations present in the average seven-year-old. One would, therefore, not expect a one-to-one correlation between volume conservation and IQ.

Nevertheless, brighter children would choose professional or academic roles more frequently than children with less ability. There would then be a tendency for bright children, more frequently than dull children, to attain the abstract conception of volume. This tendency would be reflected in a low but positive correlation of volume conservation and IQ in agreement with the correlation obtained in the present study.

In sum, the results of this investigation suggest that the appearance of formal operations and the adoption of adult roles beginning at ages 11 to 12 results in an attenuation and selectivity of inductive conceptualizing that must be taken into account before generalizing about the cognitive development of adolescents.

---

[8] For example, one girl who had an IQ of 140 but failed the volume test told me her father did not believe in women getting too much education.

SUMMARY

Four hundred and sixty-nine junior and senior high school students were tested for their conceptions of mass, weight, and volume. Results showed: (a) of the students tested, 87 per cent had attained abstract conceptions of mass and weight, but only 47 per cent had attained an abstract conception of volume; (b) the percentage of students having an abstract volume conception increased significantly between the ages of 12 and 18; (c) a significantly higher percentage of boys than girls attained an abstract conception of volume; (d) there was a low but positive correlation of IQ with attainment of the volume conception.

The results of the study were interpreted within the framework of Piaget's genetic theory of intelligence. According to this theory, the operational and structural readiness for the attainment of the volume conception appears just at the onset of formal operations and the adoption of adult roles. These two events produce an attentuation of interest and opportunities for spontaneously attaining quantity conceptions whose formation is now dependent upon the particular role the young person adopts. This conclusion was found to be consistent with the relations of age, sex, and IQ to the attainment of an abstract conception of volume.

REFERENCES

1. AUSUBEL, D. *Ego development and the personality disorders.* Grune & Stratton, 1952.
2. BALDWIN, A. L. *Behavior and development in childhood.* Dryden, 1955.
3. DENNIS, W. Animistic thinking among college and university students. *Sci. Mon.,* 1953, 76, 247-250.
4. DEUTSCHE, J. M. The development of children's concepts of causal relations. *Univer. Minn. Inst. Child Welf. Monogr.,* 1937, No. 13.
5. ELKIND, D. The development of quantitative thinking: a systematic replication of Piaget's studies. *J. genet. Psychol.,* 1961, 98, 37-46.
6. ELKIND, D. Children's discovery of the conservation of mass, weight and volume: Piaget replication study II. *J. genet Psychol.,* in press.
7. GRECO, P. & PIAGET, J. *Apprentissage et connaissance.* Paris: Presses Universitaires de France, 1959.
8. HURD, A. W. Sex differences in achievement in physical science. *J. educ. Psychol.,* 1934, 25, 70.

9. Jersild, A. T., & Tasch, R. J. *Children's interests.* Bureau of Publications, Teachers College, Columbia Univer., 1949.

10. Monash, L. Why children like or dislike school. *Understanding the Child,* 1947, 16, 67-70.

11. Piaget, J. *The child's conception of number.* London: Kegan Paul, 1952. (First published in Switzerland, 1941.)

12. Piaget, J., & Inhelder, B. *Le developpement des quantités chez l'enfant.* Neuchatel & Paris: Delachaux & Niestle, 1941.

13. Piaget, J., & Inhelder, B. *The growth of logical thinking from childhood to adolescence.* Basic Books, 1958.

14. Tenenbaum, S. Attitudes of elementary school children to school, teachers and classmates. *J. appl. Psychol.,* 1944, 28, 134-141.

STUDIES OF DISABILITY: A TECHNIQUE FOR MEASUREMENT OF PSYCHOLOGICAL EFFECTS[1]

ROBERT W. DOWNING, GEORGE MOED,*
and BYRON W. WIGHT
Children's Seashore House

This is the first report from a research project which is investigating social and psychological effects of disability and institutionalization on the growth and development of children. Disabilities can be classified according to severity, reversibility, duration, etc. Comparison of groups of children who differ with respect to such characteristics should ultimately reveal kinds and extents of psychological effects attributable to disability and institutionalization. Another objective of our research is to test parts of the major theories of disability. There are those who feel that the personality functioning of the disabled is best understood as reactions to being discriminated against (1, 3, 7), whereas others stress the importance of variations in body image (4). These ideas can be tested if appropriate techniques are available.

The concept of self-evaluation is crucial to an understanding of the role of body image, discrimination, or any other psychological process in disabled children. It seems reasonable that perception of one's functional

---

* Research Project, Children's Seashore House, Atlantic City, New Jersey.

[1] The authors are collaborators in the Research Project at Children's Seashore House, a project supported in part by a grant from the Ford Foundation and directed by Dr. George Moed. The data in this study were collected as part of the project.

The authors wish to thank their Advisory Committee members, Drs. Wilton M. Krogman, Samuel M. Wishik, and M. Erik Wright, as well as the Medical Director, Dr. Harvey N. Vandegrift and consultants, Drs. James C. Diggory, John W. Dyckman, and Seymour Feshbach, for their valuable aid and criticism.

Child Develpm., 1961, 32, 561-575.

limitations in achieving certain goals can lower self-evaluation. Equally reasonable is the idea that acceptance by the disabled of the belief that they are less worthy (because they are less capable or less attractive) than the intact can also lower self-evaluation. In either case, invidious comparisons may follow with consequent damage to the psychological well-being of disabled children and obstruction of the ordinary course of their growth and development.

This methodological report describes a semantic differential. This version of a research technique created by Osgood, Suci, and Tannenbaum (6) was constructed by us for the study of effects of disability on growth and development, self-evaluation and attitudes toward disability of disabled and intact children, and for the test of hypotheses implied by theories of disability. Reliability, factor independence, and the range of applicability of the test were investigated. Although several investigators have adapted the semantic differential technique for studies of health and disability, they have not dealt with these problems.[2]

Reliability, or consistency, of measurement is perhaps not so important if one is dealing with effects so powerful that they are obvious, or if hypotheses related to them are confirmed. If one is attempting to measure weaker influences, or if negative results are obtained, reliability becomes important. Before it can be concluded that an effect or hypothesized relationship does *not* obtain, it is necessary to ascertain reliability since an unreliable instrument can also yield negative results.

Factor independence must similarly be determined for several reasons. The test employs particular concepts designed specifically for use with disabled children, and the knowledge that the test consists of specific factors, rather than unknown components, is valuable in interpreting results. Moreover, the establishment of factor independence may serve as the foundation for testing predictions of circumstances under which independence will vanish and factors will be correlated. Fundamental correlation (nonindependence) of factors would preclude a definitive test of such predictions. The test to be described, for example, contains factors of Evaluation, Potency, and Activity, all of which are especially relevant to physical disability. An instance in which they might be predicted to be correlated would be that in disabled children inactivity leads to a decrease in feelings of potency and a low self-evaluation. This relation could hold notwithstanding the general independence of these factors.

Applicability involves a consideration of the limits of the test in terms of age and sex of subjects. It is our intention to develop this semantic differential not only for research but also as a clinical instrument amenable to qualitative interpretation on an individual basis. This may be accomplished if reliability, factor independence, and range of applicability are satisfactory.

---

[2] We wish to express our gratitude to Mrs. Judith Goldston and Drs. Cynthia Deutsch, Bernard Mausner, and Franklin Shontz for sharing their ideas and data with us.

METHOD

Test

This technique for measuring connotative meaning has been fully described by Osgood *et al.* (6). Our test employed nine scales[3] from the three major independent factors of connotative meaning extracted by Osgood *et al.* (6). The Evaluative factor was represented by the good-bad, nice-awful, and clean-dirty scales, the Potency factor by the strong-weak, large-small, and hard-soft scales, and the Activity factor by the moving-still, hot-cold, and fast-slow scales.

Research hypotheses dealing with self-evaluations, reactions to disability and relevant persons and objects, feelings toward parents and peers, and attitudes toward cultural ideals and terms of opprobrium led to the selection of the following 15 concepts:

BOY
HOSPITAL
ME
GIRL
CRUTCH
DOCTOR
WHEELCHAIR
SISSY
MOTHER
FOOTBALL PLAYER
FATHER
CRIPPLE
OLD MAID
NURSE
MOVIE STAR

For example, it was hypothesized that, as time since recovery from illness and discharge from the hospital increases, identification with the disabled decreases, identification here being measured as the difference in connotative meaning between ME and CRIPPLE. Both scales and concepts were chosen with regard for the age and limits of sophistication of the subjects. On this account, a five-point scale was used, quantified by the words "Very" and "Pretty." The format is illustrated below.

DOCTOR

(Very) (Pretty) (Pretty) (Very)

good______:______:______:______:______bad

Procedure

Subjects were instructed[4] to describe each of the 15 concepts by marking one of the spaces on each scale line and not to make more than one mark on a line. They were told to use the first space if, for example, they thought DOCTOR was very good, etc. The middle (neutral) space was to be used if they thought the scale words could not be applied to the concept. Thus, each child made 135 marks (9 scales × 15 concepts) in about 10 minutes.

---

[3] Analysis of data from this study has since led to revision of the instructions, replacement of the hot-cold scale by a sharp-dull scale, and addition of the concept NEGRO. However, this report is based on the original version.

[4] The following supplementary material has been deposited as Document number 6686 with the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C.: (a) Children's Seashore House Semantic Differential rating sheet, (b) instructions for administration of the rating sheet, (c) chance distribution of step deviation scores, (d) chance distribution of factor scores, and (e) group mean factor scores. A copy may be secured by citing the Document number and by remitting $2.50 for photoprints, or $1.75 for 35 mm. microfilm. Advance payment is required. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress.

Subjects

The data were obtained from a group of 41 children, ranging in age from 7 to 19 years, who had been hospitalized and immobilized for about two years but had since returned to the community in a relatively intact condition. This sequence of events occurs in Legg-Perthes disease, a disorder involving the femoral capital epiphysis of children (5). The group is especially interesting because they have moved from a situation of immobilization and disability to one of mobility and intactness. What are their identifications, and do their attitudes toward the disabled differ from either intact children who have never been disabled or from those of permanently disabled children? Still another interesting question concerning this group is that, if disability entails value loss (2), is recovery accompanied by value gain?

Table 1 presents data describing the subjects. The ratio of boys to girls (29:12) is a reflection of the differential frequencies with which they are affected with the disease. In addition to data on sex differences, age differences were analyzed. The total group was divided into a "Young" and an "Old" group by splitting the sample at the median and discarding the middle case. This split placed roughly half of the boys and half of the girls in each age group. Chi squares calculated for differences between the Boy-

TABLE 1

SEMANTIC DIFFERENTIAL SAMPLE CHARACTERISTICS
(All figures in months)

	Total $N=41$	*Boys* $N=29$	*Girls* $N=12$	*Young* $N=20$	*Old* $N=20$
Age at Testing					
Mean	165	162	171	133	196
Range	87–227	87–217	112–227	87–165	173–227
Length of Hospitalization					
Mean	24	24	24	21	27
Range	2– 46	2– 46	19– 41	7– 37	2– 46
Time since Discharge					
Mean	63	59	71	37	86
Range	8–107	8–107	14–106	8– 77	37–106

Girl and Young-Old subgroups in length of hospitalization were not significant. Whereas the Old group had, of course, a longer time since discharge, there was no such difference between the Boys and Girls.

Table 2 shows prorated Verbal Intelligence Quotients. Tests of differences between subgroups were not significant. The total group was slightly, but significantly, lower than the groups used in establishing norms for the Wechsler Intelligence Scale for Children and the Wechsler Adult Intelligence Scale.

TABLE 2

VERBAL INTELLIGENCE QUOTIENTS*

	Total	*Boys*	*Girls*	*Young*	*Old*
Mean	95.1	95.7	93.5	94.1	95.1
SD	14.5	12.9	17.8	15.7	12.9
Range	63–124	63–120	66–124	63–120	66–124

* Prorated Verbal Intelligence Quotients based on the Wechsler Intelligence Scale for Children and the Wechsler Adult Intelligence Scale Information, Comprehension, Similarities, and Vocabulary subtests.

Measures of Performance

Reliability. The reliability measure, similar to one described by Osgood *et al.* (6), was developed to answer the question: Are the subjects consistent in using the three scales which represent each factor?

The numbers 2, 1, 0, —1, and —2 were assigned to the points on each scale from left to right. Thus, for each concept, each subject had three scores for each factor. If a subject marked a concept at exactly the same point on the three scales representing a factor then his within-factor reliability was maximal, and the greater the scatter on these scales the lower the reliability.

For each concept the three scale markings of a factor were treated as combinations of three pairs of scores. A difference score (always considered positive) was obtained for each of the pairs. The sum of these differences is the "step deviation," or reliability score, for one subject rating one concept on one factor. If, for example, a subject obtained the following scores on the Evaluative factor for the concept ME

good-bad	2
clean-dirty	1
nice-awful	0

then his step deviation score was $(2 - 1) + (2 - 0) + (1 - 0) = 4$. With this method of scoring the only possible step deviation scores are 0, 2, 4, 6, and 8. Zero step deviations come from three identical scale ratings while scores of 2 require two identical ratings with a third only one position away.

It seems reasonable to assume that scores greater than 2 step deviations reflect unreliability of the subject, or the scales, or a significant scale-concept interaction. Since only two kinds of combinations yield scores of 0 or 2, the reliability of these scores is unambiguous and their occurrence by chance alone is small. However, many combinations can generate scores of 4, 6, and 8 so that the meaning of these scores is not known to us and they can more readily be attributable to chance. Setting the reliability criterion at scores of 0 and 2 means that no subject's ratings can be counted reliable if they lie on both sides of the neutral space, the implication being that such ratings reflect a conflict of feeling about a concept. In assessing reliability, the frequencies of obtained step deviation scores of 0 and 2 were compared with those expected by chance. Chance frequencies were determined by generating the distribution of all possible combinations of three scale scores and counting the numbers of ways of obtaining each of the possible step deviation scores (0, 2, 4, 6, and 8). Scores of 0 and 2 are expected, on the average, 23 per cent of the time by chance alone.[4]

Step deviation scores of 0 and 2 were summed for each concept on each factor separately, yielding three distributions of reliability scores in which the 15 concepts were the N. If the factors are reliable, then frequencies of 0 and 2 scores in each one of these distributions should be significantly large. If reliabilities of the factors are independent, then correlations between pairs of them should be nonsignificant. If the instrument has a wide applicability, then tests of differences, on this index, between either age or sex subgroups should be nonsignificant.

Factor scores. Just as the reliability score represents variability of the scale scores, factor scores represent their central tendency. For each concept a subject's scores on the three scales representing a factor were averaged to obtain his factor score. Following the illustration above where a subject marked the Evaluative scales 2, 1, and 0 for the concept ME, his factor score was $(2 + 1 + 0) \div 3 = 1$. Factor scores derived from the five-point scale can range from $+2$ to -2. Scores as large as 2 can be obtained only from three identical scale scores of 2. The probabilities of obtaining them by chance alone are much smaller than the probabilities of obtaining smaller factor scores (like 1 or 0) since these may be arrived at through many different combinations of three scale scores. When it was desired to compare obtained factor scores with those expected by chance, these differential probabilities proved valuable since they are finite, exact, and calculated.[4]

The psychological significance of these factors lies in their relevance for describing the self-perceptions of disabled children and their ratings of important persons and objects. In the test they are used to index the amounts and directions of evaluation, potency, and activity attributed to concepts. One might hypothesize that children who are severely disabled would attribute a higher evaluation to concepts like DOCTOR and NURSE (upon whom they are quite dependent) than children suffering less incapacitation. At a more descriptive level it might be interesting to compare disabled children's self-ratings of potency with those of intact children.

Group mean factor scores were derived for each concept on each factor separately, yielding three distributions of group factor scores in which the 15 concepts were the *N*. If the factors are independent, then correlations between pairs of these distributions should be nonsignificant. If the instrument has wide applicability, then tests of differences, on this measure, between either age or sex subgroups should be nonsignificant.

Meaning. Geometrically, the E, P, and A factor scores are the coordinates, in a three dimensional semantic space, of a point representing a concept for a subject. The point where these dimensions bisect each other and form perpendicular axes represents zero, or the origin, in the space. Meaning, as a point in space, can be composed of positive as well as negative factor scores in addition to magnitudes of these scores. Thus, it is possible to characterize concepts and to compare characterizations derived from responses of individuals and groups. The meanings attributed to concepts concerning disability can be better understood by measuring their relative

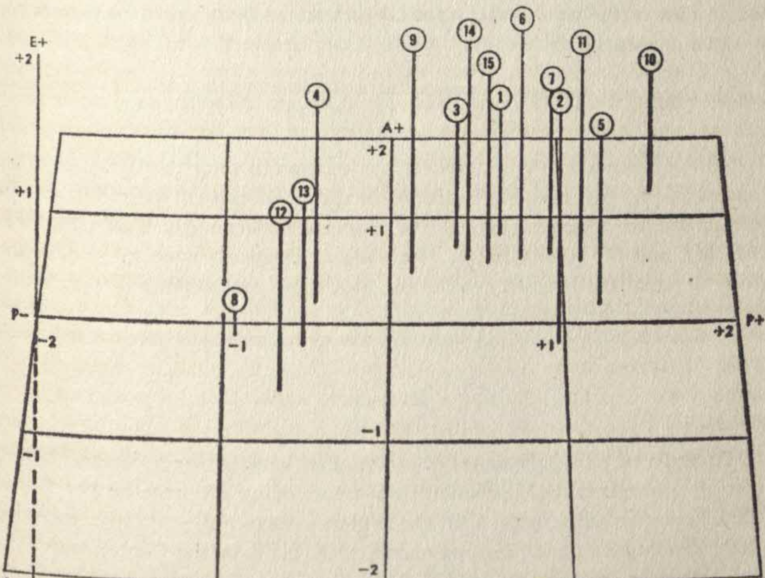

FIGURE 1—Semantic space for former Legg Perthes patients. Concepts: (1) boy, (2) hospital, (3) me, (4) girl, (5) crutch, (6) doctor, (7) wheelchair, (8) sissy, (9) mother, (10) football player, (11) father, (12) cripple, (13) old maid, (14) nurse, (15) movie star.

proximity to concepts which have generally stereotyped connotations, such as the ideals and terms of opprobrium included in the test. The concept FOOTBALL PLAYER should be characterized positively while SISSY should be characterized negatively. Will CRIPPLE be characterized similarly to FOOTBALL PLAYER, SISSY, or neither? Discernment of relations among the characterizations of concepts could be used in personality investigation and in the test of theories of disability.

One approach to the analysis of meaning was to construct a three dimensional model of the semantic space and mark the locations of the concepts. This model (*see* Figure 1) facilitated the interpretation of each concept's characterization as well as the relations between characterizations. The limits of applicability of the test for the age and sex subgroups were explored by testing for the significance of differences in meaning between subgroups on each concept. Finally, the differences between obtained meanings for each concept and those expected by chance were calculated to test the appropriateness of the concepts.[4]

Distance. Osgood *et al.* (6) have detailed the method of investigating distances between concepts. Just as any concept can be compared with a zero point, or origin, in a semantic space, so can distances between concepts be compared. As expressed in the formula below, distance is defined as the square root of the sum of the squared differences between factor scores for any two concepts.

$$D_{12} = \sqrt{(E_1 - E_2)^2 + (P_1 - P_2)^2 + (A_1 - A_2)^2}$$

Self-perceptions can be interpreted in relation to their proximity to other concepts even apart from the locations of these concepts in space. Of obvious interest are such comparisons as the distance between ME and CRIPPLE with ME and BOY (or GIRL). The distance approach is an avenue for the study of identification, self-evaluation, and some components of personality functioning of the disabled.

RESULTS

Reliability

Analysis of reliability scores of 0 and 2 by concepts over factors showed that all concepts exceeded chance frequencies of 23 per cent on the Evaluative factor in all groups. On the Potency factor all concepts exceeded chance frequencies in all groups except MOTHER in the Girls group. The only instances of concepts failing to exceed chance on the Activity factor were HOSPITAL in the Girls and Young groups and WHEELCHAIR in the Old group. The results also showed that reliable responses were fairly evenly distributed among the concepts.

Chi square analysis of reliability scores of 0 and 2 by subjects over factors showed that significant numbers of subjects exceeded the 0 and 2 criterion

TABLE 3

NUMBERS OF SUBJECTS WHO EXCEEDED THE CHANCE FREQUENCY OF 0 AND 2 RELIABILITY SCORES

	GROUP				
Factor	*Total*	*Boys*	*Girls*	*Young*	*Old*
E	41	29	12	20	20
P	39	28	11	18	20
A	31	19	12	13	17

on each factor (*see* Table 3). Table 4 shows 0 and 2 scores as percentages of all scores; with the exception of the Young group on the Activity factor all percentages were significantly large.

Analysis by concepts, subjects, or responses pointed to the Activity factor as the least reliable, and examination of this factor showed that the lower reliability was attributable to a large number of neutral responses on the hot-cold scale (*see* Table 5). Evidently, this scale was inappropriate for some concepts and subjects. (In order to achieve greater reliability, the hot-cold scale has since been replaced by a sharp-dull scale in a revised form of the test.)

Although the Evaluative, Potency, and Activity factors all showed reliability significantly greater than chance, they differed considerably on this measure ranging from high to low in the order mentioned. In Table 6 the between-factor rank order correlations of reliability scores for all groups were not significant, reflecting factor independence as measured by 0 and 2 scores.

Table 7 provides data concerning reliability score differences in the age and sex subgroups. The Wilcoxon matched-pairs signed-ranks test was used to evaluate these differences. On the Potency and Activity factors the Old

TABLE 4

0 AND 2 RELIABILITY SCORES PRODUCED BY EACH GROUP ON EACH FACTOR
(All figures expressed in percentages)

	GROUP				
Factor	*Total*	*Boys*	*Girls*	*Young*	*Old*
E	80	81	79	77	83
P	53	53	52	46	59
A	42	42	43	35	47

TABLE 5

NEUTRAL RESPONSES BY SCALES AND FACTORS

Scales	*N*	%
good-bad	43	3
nice-awful	63	4
clean-dirty	46	3
Σ Evaluative	152	10
strong-weak	66	5
hard-soft	208	15
large-small	212	15
Σ Potency	486	35
moving-still	160	11
hot-cold	450	32
fast-slow	172	12
Σ Activity	782	55
Total	1420	100

group was clearly more reliable than the Young group. This higher reliability, not attributable to sample size or IQ measures, probably reflects the greater cognitive and affective differentiation of the older children. As can be seen from Tables 3 and 4, lower reliability patterns came from the Young group on all factors but especially from the Activity factor. Interpretation of responses from children much younger than this sample or from the Activity factor alone should be made with caution.

Thus far, the problems of reliability, factor independence, and general applicability of the test have been examined by using step deviation scores of 0 and 2. The results, by and large, confirmed the independence of the factors and established the reliability and applicability of the test with certain limitations relating to age and to the Activity factor.

TABLE 6

BETWEEN-FACTOR CORRELATIONS OF RELIABILITY

Factor Comparisons	*Total*	*Boys*	*Girls*	*Young*	*Old*
E - P	—.20	—.14	—.19	—.10	—.01
E - A	.23	.36	.07	—.18	.33
P - A	—.21	—.11	—.34	—.16	—.19

TABLE 7

BETWEEN-GROUP DIFFERENCES IN RELIABILITY

	SMALLER SUM OF RANKS	
Factor	*Boys-Girls*	*Young-Old*
E	40.0	16.5
P	58.5	11.0*
A	49.5	16.5*

* $p < .05$.

Factor, Meaning, and Distance Scores

Because of the interrelations of factor, meaning, and distance scores, they are treated together in the results which follow. A preliminary part of the analysis was undertaken to determine appropriateness of the concepts by comparing the obtained distributions of factor scores for each concept with the distribution expected by chance. Forty-five comparisons (15 concepts on three factors) were made using the Kolmogorov-Smirnov one-sample test (*see* Table 8). There were 39 factor distributions different from

TABLE 8

MAXIMUM PERCENTAGE DEVIATION FROM CHANCE OF OBTAINED FACTOR SCORE DISTRIBUTIONS

Concept	*Evaluation*	*Potency*	*Activity*
Boy	.41*	.43*	.52*
Hospital	.85*	.45*	.16
Me	.62*	.26*	.43*
Girl	.77*	.29*	.20
Crutch	.67*	.67*	.28*
Doctor	.85*	.40*	.38*
Wheel Chair	.67*	.52*	.40*
Sissy	.09	.52*	.20
Mother	.82*	.13	.30*
Football Player	.43*	.72*	.70*
Father	.79*	.58*	.38*
Cripple	.62*	.36*	.36*
Old Maid	.57*	.31*	.18
Nurse	.84*	.33*	.40*
Movie Star	.74*	.33*	.40*

* $p < .05$.

chance at the .05 level, indicating a high degree of agreement among subjects as to the connotative meaning of these concepts. The six nonsignificant results were: SISSY on E; MOTHER on P; HOSPITAL, GIRL, SISSY, and OLD MAID on A. The possibility that these results might be due to offsetting directions of meaning attributed to the concepts by the subgroups was examined and rejected except for the one instance of HOSPITAL on the A factor. The extent to which these distributions are due to ambivalence of meaning induced in individual subjects by these concepts requires further study.

Further analysis of factor scores was undertaken to determine the agreement of the subgroups on connotative meanings attributed to the concepts. Differences in factor score distributions were tested for each concept with the Kolmogorov-Smirnov two-sample test (*see* Table 9), and chi squares for the three factors were summed inasmuch as reliabilities (errors of measurement) of the factors have been shown to be independent.

In only one instance was the *combined* chi square significant at the .05 level, indicating a significant distance between subgroups. Boys and Girls disagreed on the characterization of the concept GIRL with Girls attributing greater potency and activity to this concept. There were several additional instances where significant differences were found on *individual* factors. Girls rated CRUTCH as more active. The Old group had a more positive evaluation of WHEELCHAIR and OLD MAID and attributed greater potency to ME and more activity to HOSPITAL than the Young group.

TABLE 9

BETWEEN-GROUP DIFFERENCES IN MEANING

	COMBINED CHI SQUARES	
Concept	*Boy-Girl*	*Young-Old*
Boy	3.00	8.80
Hospital	5.25	10.10
Me	4.29	10.10
Girl	15.00*	6.70
Crutch	10.02	3.80
Doctor	8.04	2.20
Wheel Chair	5.07	9.30
Sissy	4.62	11.00
Mother	7.35	5.70
Football Player	2.47	4.90
Father	7.57	1.70
Cripple	2.49	4.30
Old Maid	5.40	10.90
Nurse	6.91	5.40
Movie Star	6.85	2.20

* $p < .05$.

Further evidence for the applicability of the test may be seen in the correlations of factor scores in Table 10. Subgroups generated similar orders of concepts. However, the Girls rated more positively than the Boys on all factors and most concepts. This may reflect the operation of a cultural norm which permits greater expression of affect on the part of females. The small number of significant differences and the high correlations between subgroups in the rating of concepts means that, by and large, subgroups can be treated together in the testing of hypotheses and that separate norms are not required for interpretation of results. General applicability of the test was thus supported.

TABLE 10

BETWEEN-GROUP CORRELATIONS OF FACTOR SCORES

Factor	*Boys-Girls*	*Young-Old*
E	.91*	.85*
P	.98*	.93*
A	.91*	.92*

* $p < .05$.

Unique characteristics of the subjects and the concepts demanded a test of factor independence. Rank order correlations between group mean factor scores were calculated (*see* Table 11). The E factor was independent in all comparisons. Although only one correlation was significant, those in the P-A comparisons were higher than the others. In general, it may be said that factor independence was sustained with these peculiar characteristics of concepts and subjects.

TABLE 11

FACTOR INTERCORRELATIONS

Factor Comparisons	*Total*	*Boys*	*Girls*	*Young*	*Old*
E - P	.23	.23	.16	.29	.15
E - A	—.02	—.17	.20	—.06	—.07
P - A	.49	.33	.48	.37	.54*

* $p < .05$.

Figure 1 shows the locations of the concepts as characterized by the total group. None of the concepts had negative E scores, although one was close to zero. Perhaps the children reflected a cultural prohibition against evaluating objects or persons negatively in the presence of adults (the experimenters). The most negatively characterized concept was SISSY, and

RESPONSE SETS AND MANIFEST ANXIETY SCORES IN A RETARDED POPULATION

CHARLES G. MATTHEWS*
Fort Wayne State School

and LEON H. LEVY
Indiana University

The prominence of the construct of anxiety in the thinking of psychologists demands that any instrument purporting to yield a measure of it be subjected to the closest possible scrutiny. The present study has as its purpose, therefore, an examination of the applicability of the Children's Form of the Manifest Anxiety Scale (CMAS) (1) for use with mentally retarded population. Since the CMAS was standardized on normal children in the fourth, fifth, and sixth grades and since it has been shown to be related to intelligence and school achievement (2, 6), one might question the meaning of scores for mental defectives on this scale. In the present study this problem was investigated in several ways. The first was by an examination of the test performance per se of retardates, i.e., their group mean and standard deviation, test-retest reliability, and item analysis data; the second, more directly aimed at the construct validity of the CMAS when used with mental defectives, involved the correlation of CMAS scores with performance on a digit span recall test based on the hypothesis that anxiety interferes with the abilities tapped by the digit span test. That is, from a variety of different theoretical positions, anxiety has been conceptualized in such a way that high anxiety would be expected to interfere with learning and performance. Thus, if the CMAS is a valid measure of the construct of anxiety, scores on it should correlate negatively with a digit span recall test. Although the evidence for the validity of this hypothesis is far from univocal (4, 5, 7, 8), finding a significant negative correlation could be taken as supporting evi-

---

* Department of Psychological Services, Fort Wayne State School, Fort Wayne, Indiana.

Child Develpm., 1961, 32, 577-584.

dence for the construct validity of the CMAS when used with a retarded population.

Still related to the meaning of CMAS scores for a retarded population, but possibly having wider implications, is the question of the contribution of response sets. Because items only contribute to the anxiety score if answered "Yes," it seems reasonable to hypothesize that acquiescent or assent tendencies contribute significantly to the score obtained, this, all the more so, for a retarded population, because of the common observation of the greater acquiescence and persuasability of this group. Since, with the exception of two items, the 11 items constituting the Lie Scale of the CMAS contribute to the L Score when answered "Yes," we might expect acquiescence to affect this score as well. Thus, the third approach to the meaning of CMAS scores for retardates involved the correlation of these scores with a specially constructed instrument to measure assent and dissent tendencies. From the standpoint of the response set hypothesis, the predictions were made that both the anxiety score and the lie score would correlate positively with the measure of assent tendencies and negatively with the measure of dissent tendencies.

METHOD

Subjects

Thirty mildly retarded male residents of the Fort Wayne State School were used in this investigation. Etiological background in the experimental group was heterogeneous, with diagnostic labels of "familial" and "congenital cerebral maldevelopment—nonspecific" contributing the bulk of the cases. Patients having gross visual, auditory, or motor handicaps were excluded from the sample, as were residents with histories of psychotic episodes or other marked affective or ideational disturbances.

Subjects ranged in age from 18 to 30 years. IQ scores ranged from 48 to 72, with scores based upon 24 Full Scale WAIS tests and six Stanford-Binet (L) results obtained within two years of the time of this investigation. While it is not strictly defensible to combine scores from different psychometric tests to yield a group mean, it was felt that such a procedure would serve to describe in a general way the intellectual characteristics of the group (IQ mean = 59.5; SD = 6.7).

Procedure

Subjects were seen individually, with uniform experimental setting and procedure for each subject. Immediately prior to test administration, each subject was queried briefly about cottage placement and about recreational interests and activities in an effort to convey an attitude of informality and acceptance to the subject. Following this brief interview, the CMAS was administered to the subject. Because of the nature of the experimental population, each item in the CMAS was read to the subject and his responses recorded. CMAS instructions were given as follows:

> I have a number of sentences or statements on this paper that I am going to read to you. I want you to listen carefully to each one I read, and to tell me whether the sentence says something that is true about you or not true. If the sentence I read says something that describes you or the way you feel, say "True"; if the sentence is not true or does not describe you, say "Not True."

Upon completion of the CMAS, the Digit Span subtest from the Wechsler-Bellevue I was administered to each subject, and highest number of digits recalled was recorded. In order to obtain a measure of response stability on the CMAS, the scale was re-administered to each subject two weeks after the first administration. The departure in administration of the CMAS in the present study from that used in its standardization was required for obvious reasons, and its significance in accounting for the subsequent difference found cannot be gainsaid without further empirical study. On the other hand, it might be contended that if the CMAS were to be used with a retarded population it could not be administered in any other way and so must be evaluated on these grounds.

In order to investigate possible relations between response set, CMAS scores, and Digit Span performances, the Assent-Dissent Response Set Scale (ADRSS) was developed by the authors and administered to each subject immediately following the second administration of the CMAS.

The ADRSS consists of 16 statements which are read to the subject by the examiner. The subject is requested to answer "True" or "Not True" to each statement. Eight pairs of contradictory statements are used in the ADRSS, with the statements in each pair sequentially interrupted by statements from other pairs used in the test. For example, statement 1 is: "My left arm is a little longer than my right arm"; its contradictory statement is number 10: "My arms are both the same length." Similarly, statement 3 is: "I like watching TV more than playing sports"; while statement 11 is: "I like sports better than TV."

Scoring is based upon the number of contradictory pairs in which both statements are answered in a particular direction. Since the subject can answer either "True" or "Not True," the ADRSS yields both an assent response set score and a dissent response set score. Assent response set scores may range from 0 to 8 and are based upon the number of contradictory pairs in which both statements are answered "True." Dissent response set scores may range from 0 to 8 and are based upon the number of contradictory pairs in which both statements are answered "Not True."

RESULTS

Test Performance

Table 1 presents the test and retest means and standard deviations for the anxiety and lie scores as well as their test-retest reliabilities. For comparison purposes, similar data are presented for the 73 sixth grade boys of

TABLE 1

TEST PERFORMANCE OF RETARDATES AND SIXTH GRADE BOYS

	TEST ($N = 30$)		RETEST ($N = 30$)			6TH GRADE BOYS ($N = 73$)		
	Mean	*SD*	*Mean*	*SD*	*r*	*Mean*	*SD*	*r*
Anxiety Score	22.2	8.7	19.0	9.4	.84	16.6	7.4	.82
Lie Score	5.1	2.3	4.4	2.7	.86	1.8	1.6	.53

the standardization group (1), the group probably closest to the present one in mental age. Here it will be noted that, for both the anxiety and lie scores, the means and standard deviations of the retarded group are larger. Comparing the first test mean anxiety and lie scores of the retardates with those of the sixth graders by means of the *t* test revealed differences significant at the .01 and .001 levels, respectively. The variances for the two groups for these two scores, when compared by means of the *F* ratio, were found to be significantly different ($p < .05$) in the case of the lie score, thereby making the *t* test for the lie score means only somewhat suspect as to meaning. Comparing test-retest reliabilities for the two groups—a comparison not entirely justified since the test-retest interval for the sixth grade boys was only one week—revealed no difference for the anxiety score, but a difference significant at the .001 level for the lie score, with the retardates showing the greater stability.

Finally, the correlation between first test mean anxiety and lie scores for the retardates was $-.04$. This compares favorably with the corresponding correlation of $-.10$ obtained for the standardization group, suggesting that in at least one aspect of their performance the two groups do not differ significantly.

To get some indication of population item-response characteristics, an item analysis was performed on the anxiety items, counting the number of "True" responses for each item for the entire group. In Table 2 are presented those items answered "True" by 80 per cent or more, or by 20 per cent or less, of the total group.

The items answered "True" by 80 per cent or more of the group, it will be noted, represent concerns which might be expected and considered appropriate for a person removed from his home and placed in an institution (items 8 and 46) and of limited ability (items 9, 29, and 44). Furthermore, of 11 items having some reference to ability, achievement, or attention span, nine were answered "True" by 50 per cent or more of the group. For these same items, Hafner and Kaplan (3) found only four on which 50 per cent or more of their 62 normal fifth grade boys answered "Yes." These data, therefore, strongly suggest that the item content of the CMAS is such as to discriminate between children of differing ability levels and possibly

TABLE 2

CMAS ANXIETY ITEMS ANSWERED "TRUE" BY 80 PER CENT OR MORE, OR BY 20 PER CENT OR LESS OF THE TOTAL GROUP

Item Number	*Content*	*Per Cent Answering "True"*
29	I worry about doing the right things	93
8	I wish I could be very far from here	90
46	I often worry about what could happen to my parents	90
44	I often do things I wish I had never done	83
9	Others seem to do things easier than I can	80
19	Often I have trouble getting my breath	20
23	I have to go to the toilet more than most people	20
26	I have trouble swallowing	03
39	I am afraid of the dark	00

also between institutionalized and noninstitutionalized children. The significant difference in mean anxiety scores, already noted, might further be taken as confirming this.

Of the four items answered "True" by 20 per cent or less of the retardates, three are concerned with somatic functioning. For these same items, Hafner and Kaplan found frequencies of 16, 7 and 5 per cent, suggesting that, if somatic complaints are highly correlated with anxiety, the groups may not be as different as their CMAS scores would suggest. More simply, perhaps, the similarity between normals and retardates in responses to the somatic complaint items may indicate that this is an area which is little affected by either intellectual ability or institutionalization.

Evidence for Construct Validity

The correlation between the anxiety score and digit span recall was .00. The corresponding correlation for the lie score was —.04. Thus, to the extent that the hypothesis linking anxiety with interference with performance on such tasks as digit span is valid, the present findings must be counted as negative evidence for the construct validity of the CMAS when used with retardates.

Evidence for Response Sets

The mean assent score on the ADRSS was 1.5 with a range of 0 to 6, while the mean dissent score was .77 with a range of 0 to 3. The correlation between assent and dissent scores was —.40, significant at the .02 level by a one-tailed test, justified in this instance by the logic of the scale which demands that the correlation be negative for the scale to have the meaning imputed to it. The correlations of assent and dissent scores with the anxiety score were +.31 and —.37, respectively, both significant beyond the .05 level by a one-tailed criterion, thus supporting the response set hypothesis in connection with the anxiety score. The corresponding correlations with the lie score were +.31 and +.33, the former significant at the .05 level by a one-tailed criterion and the latter correlation being in the wrong direction. Therefore, the response set hypothesis is only partially supported in the case of the lie score.

Because it might be contended that the ADRSS correlations with the CMAS merely reflect a relation between assent and dissent tendencies and anxiety, the two ADRSS scores were correlated with digit span recall. If this reasoning were correct, assent would correlate negatively with digit span and dissent positively. The obtained correlations were —.34 and +.09, respectively, both in the proper direction for support of the above reasoning but with only the former correlation being significant at the .05 level by a one-tailed criterion. However, this finding and the reasoning behind it loses its cogency when it is remembered that the CMAS itself failed to correlate at all with digit span recall. Therefore, unless we are willing to conceive of the assent score as a better indicant of anxiety than the CMAS, there is no reason be believe that the correlations between ADRSS scores and CMAS scores represent anything other than the contribution of response sets to test score variance.

Discussion

While CMAS scores of mental defectives appear to be as stable as those of normals, the results of this study point fairly conclusively to the inapplicability of the test as a measure of anxiety in mental defectives. Put more conservatively perhaps, whatever it is that the CMAS score represents in a normal population, it is evident that it represents something different for a retarded one. Not only would the CMAS be likely to portray retardates as more anxious than normals, but also as less truthful. Whether either of these is correct or not cannot be ascertained by an instrument composed of items such as those contained in the CMAS which are so obviously situationally and intellectually biased or by an instrument as susceptible to response sets as the CMAS appears to be.

The absence of correlation between digit span recall and CMAS cannot be considered too damaging and may be considered as reflecting as much upon the validity of the theory leading to the prediction that such a correlation would exist as it does upon the construct validity of the test itself.

On the other hand, the support found for the response set hypothesis in the present study points beyond the problem of the meaning of CMAS scores for retardates. For recent research demonstrates that response sets may be readily found among normals as well and that a test such as the CMAS keyed only in one direction should be considered suspect on a priori grounds alone. By constructing a separate scale to measure response sets in the present study, rather than following the usual practice of simply reversing items in the test under study, it was believed that we would be tapping a more general disposition. Because of this, it would seem that the correlations obtained between ADRSS and CMAS scores are all the more significant, for these suggest that the response sets measured are not test specific. It is quite likely that, had the usual procedure of reversing items been followed or had the ADRSS been somewhat longer, thereby permitting a greater range of talent to be manifested, the correlations obtained would have been somewhat larger.

Although evidence has been presented for the operation of response sets in the CMAS, it is evident from test performance alone that they do not account for all of the variance. For one thing, if they did, the correlation between the anxiety and lie scores would be much higher than it is. For another, the item analysis would have revealed greater skewing of items in terms of frequency of "True" answers. Thus, of the 42 anxiety items, 26 were answered "True" by 50 per cent or more of the group; of the 11 lie items, six were answered "True" by 50 per cent or more of the group, thereby indicating in both instances only some skewing in the direction expected if an assent response set were operating.

Thus, it may be concluded that, for mental defectives, response sets account for some of the variance in the CMAS, but that a major portion of it is probably also determined by such factors as intellectual status and institutionalization. Whether any significant portion of CMAS score variance remains which may be meaningfully defined as anxiety remains uncertain.

SUMMARY

In an attempt to assess the applicability of the Children's Form of the Manifest Anxiety Scale to a mentally retarded population, the test was administered to 30 institutionalized, mentally retarded males on a test-retest basis together with a digit span test and a specially constructed response set scale. CMAS scores were found to be stable over time, but significantly higher than those for the sixth grade boys used in its standardization. No correlation was found between CMAS and digit span, but significant correlations were found with the response set scale. On the basis of these data, as well as an item analysis, it was concluded that the scores of retardates on the CMAS are strongly influenced by response sets, as well as situational and capacity variables, and that the test, therefore, is not appropriate for a retarded, institutionalized population.

REFERENCES

1. Castaneda, A., McCandless, B. R., & Palermo, D. S. The children's form of the manifest anxiety scale. *Child Develpm.,* 1956, 27, 317-326.

2. Hafner, A. J., & Kaplan, A. M. Children's manifest anxiety and intelligence. *Child Develpm.,* 1959, 30, 269-271.

3. Hafner, A. J., & Kaplan, A. M. An item analysis of the children's manifest anxiety scale. *Child Develpm.,* 1959, 30, 481-488.

4. Keller, J. E. The relationship of auditory memory span to learning ability in high grade mentally retarded boys. *Amer. J. ment. Def.,* 1957, 61, 574-580.

5. Lewinski, R. J. The psychometric pattern: I. Anxiety neurosis. *J. clin. Psychol.,* 1945, 1, 214-221.

6. McCandless, B. R., & Castaneda, A. Anxiety in children, school achievement and intelligence. *Child Develpm.,* 1956, 27, 379-382.

7. Moldowsky, S., & Moldowsky, P. Digit span as an anxiety indicator. *J. consult. Psychol.,* 1952, 16, 115-118.

8. Shoben, E. J. The Wechsler Bellevue in the detection of anxiety; a test of the Raskis-Welsh hypothesis. *J. consult. Psychol.,* 1950, 14, 40-45.

RELATIONS BETWEEN NURSERY SCHOOL ATTENDANCE AND TEACHERS' RATINGS OF SOME ASPECTS OF CHILDREN'S ADJUSTMENT IN KINDERGARTEN[1]

Ann Wilson Brown *and* Raymond G. Hunt*

Washington University, St. Louis

Perhaps one of the more significant aspects of change in western culture has to do with the increasingly prominent role played by educational institutions in child development. Progressing from relatively exclusive concern with fostering the literacy of its charges, the school has assumed primary responsibility for vocational preparation and, in contemporary concepts, has evolved as one of the salient cultural agents in the general socialization process.

Accompanying these shifts in the focus of educational concern has been an equally notable extension in time of institutional responsibility. On the one hand, secondary and higher educational activities have extended farther into young adulthood (and even beyond), while at the same time a "downward" extension of educational responsibility has steadily formed in the establishment of the kindergarten and other "preschool" programs.

All of this can undoubtedly be correlated with generalized mutations in western culture and particularly with changing concepts of the status and role of the family (11). In any event, it seems quite clear that for some time to come educational institutions will play an ever growing part in the socialization process.

---

* Division of Child Psychology, Washington University School of Medicine, St. Louis 10, Missouri.

[1] This paper is based upon an M.A. thesis submitted by AWB to the Graduate Board of Washington University.

Child Develpm., 1961, 32, 585-596.

For these reasons it is highly desirable that empirical study of relations between the operations of such institutions and child development be pursued with vigor. Accordingly, the present study directed itself to an investigation of relations between some aspects of children's adjustment in kindergarten and their prior attendance at a "progressive" nursery school—a problem area wherein previous researches, using a wide variety of procedures, have yielded results which are, as a body, equivocal to say the least (1, 3, 4, 5, 10, 13, 14, 19, 21).

METHOD

Design of the Study

The basic structure of the study followed a paired replicates model. Each pair consisted of two children drawn from the same kindergarten class, one of whom had attended nursery school for at least one year and the other of whom had not. Each pair was matched for social status, sex, ordinal position in the family, and IQ.

Social status was determined according to Hollingshead's two factor index (12). This index yields five status levels based upon both the education and the occupation of the principal wage earner, occupation being weighted more heavily.

Ordinal position matching was in terms of "only child," "oldest child," "youngest child," or "intermediate position." It should be noted that, in view of the relatively small size of nursery school families, matching for ordinal position provides a partial control for differences in family size. IQ was determined by administering individually to each child the Ammons Picture Vocabulary Test, Form A (2).

Finally, it should be pointed out that each pair was closely equated for age automatically as a result of grade and IQ matching.

Subjects

A total of 84 children served in this study. Half of these had attended the Washington University Nursery School during the year preceding the study (1957-1958). There were a total of 46 children in this nursery school class, but three had moved out of town and a fourth child's nonparticipation will be discussed later. All were in kindergarten when the study was made.

The children attended a total of 17 public and two private schools, chiefly in suburban communities. They were children mostly of upper-middle and lower-upper class families. Using Hollingshead's index, 64 per cent of the children fell in the highest class level, 21.5 per cent in the second, 12 per cent in the third, 2.5 per cent in the fourth, and none in the lowest. The skewed distribution obtained was, of course, a consequence of the characteristics of the nursery families, most of which were in the higher status ranges. Family sizes ranged from one to five children with a mean of 2.9. Of the 84 children who served as *Ss*, 60 were boys. The youngest child was 5 years, 4 months of age, and the oldest was 6 years, 6 months. IQs ranged from 97 to 146.

Materials

Four independent graphic rating scales were constructed according to criteria suggested by Guilford (9). Each of these was separately mimeographed on 8½ by 11 inch sheets and consisted of a continuous vertical line five inches long with points marked "very good," "good," "fair," and "poor" spaced evenly along it. Equal space was left between the lowest point marked ("poor") and the origin of the scale as well as between the highest point ("very good") and the terminus.

This procedure allows the rater to choose any point along the scale which he deems appropriate (the marked points representing orientation guides) and was thus consistent with a conception of adjustment as evaluatively determined and continuous. It has the additional merit of "spreading" the ratings, thus helping to control any "central tendency" error.

The four scales produced in this manner constitute the adjustment scales. The first of these (the "activities" scale) related to the child's adjustment to kindergarten activities. It was intended to tap the general level of adjustment exhibited by the child in his routine attitudes toward and participation in usual classroom activities. A second or "group" scale had to do with the child's adjustment to his peers as this was manifest in the classroom and as perceived by his teacher. The scale referred to the general level of adjustment exhibited in the child's relations with the other children in his classroom. The third scale ("authority") had to do with the general level of adjustment exhibited by the child in his routine relations with his teacher. Finally, the fourth scale ("personal") referred to the child's personal or "inner" adjustment. It directed the rater to consider both her knowledge of the child within the classroom and outside of it and to think in terms of the general level of adjustment felt to characterize the child "as a person." These four scales were construed as dealing with at least partially independent facets indicative of general social adjustment. They reflect what may be thought of as four primary aspects of the child's general or over-all school adjustment.

A fifth rating, of relative brightness (intellectual), was obtained via a forced choice procedure. To obtain this rating, names of paired children were listed and the rater asked simply to encircle the name of that child she felt to be "brighter." The rater was instructed always to choose one, even if the two seemed about equally bright. This scale was also mimeographed on a separate 8½ by 11 inch sheet and included, in addition to the names of the critical pairs, paired names of other children selected at random from the same class. In this way it was hoped to exercise a small measure of control over possible biasing effects attributable to the rater's guessing the purpose of the ratings. The five rating scales were bound together with a face sheet containing instructions to form a six-page booklet.

Procedure

Information concerning the major wage earner's occupation and education, number of siblings, school attended, and other pertinent information was obtained by direct telephone contact with the parents of the nursery children. The same information for the control children was obtained from the records of the particular school the child was attending.

After a match had been found for a particular nursery child, the teacher was provided with a rating booklet and asked to fill it out. She remained ignorant of the purpose of the ratings until after she had completed them. The rater was left relatively free to employ whatever criteria she deemed relevant in making her ratings—only general orienting suggestions were made. Twenty-three teachers were involved in this study.[2]

It was not possible in all cases to find a perfect match for each nursery child from among children in the same class. Since the primary dimension for matching involved the use of a classmate who had not attended nursery school, one nursery child had to be discarded from the sample owing to the absence of any child from the class who had not attended an organized nursery school.

Children who had attended parent cooperatives were included in the study as control children since this type of organization is not designed to "teach" as is the nursery school. In a few cases it was impossible to match ordinal position exactly so it was as closely equated as possible, that is, to the next category. IQ was matched within 10 points except for one exceptionally high child (IQ = 146) whose closest equal was 18 points lower. Social status and sex were matched in all cases. Generally, each pair was well matched in all respects.

RESULTS

The Adjustment Ratings

After the various ratings had been obtained according to the procedure outlined above, each of the four adjustment scales was divided into 10 equal intervals and quantified by assigning to each interval a numerical value ranging from 1 for the lowest interval to 10 for the highest. In this way each rating could be scored from 1 to 10, with a score of 1 indicating the poorest level of adjustment and a score of 10 indicating the highest level. Actually (as might be expected) there were no ratings in either of the highest or lowest intervals. The adjustment ratings on each scale ranged from 2 through 9 inclusive, with a modal rating of 6. Some skewness toward the low end of the scale was evident in the distribution of individual ratings.

---

[2] Unfortunately, it was not possible to obtain direct estimates of the reliability of the teachers' ratings. However, it would seem a reasonably safe assumption that reliability would be high as all of the teachers were rating children well known to them and along dimensions not at all alien to their thinking about these children. Furthermore, since a relatively large number of teachers was used, it is likely that personal peculiarities of any particular teacher which might attenuate the reliability of her ratings would be well compensated in the over-all results. Finally, since the ratings were obtained late in the school year, it is unlikely that marked changes would occur in the children themselves sufficient to alter the results presented. It might be added that in other similar (though as yet unpublished) studies these same scales have produced highly reliable ratings from preschool teachers.

Differences in rated adjustment between the nursery and control children were evaluated by performing independent t tests of the difference between the mean adjustment scores for each group on each of the four adjustment scales. The results of these analyses, which followed a procedure for paired observations suggested by Dixon and Massey (6), are reported in Table 1. In this table N represents number of pairs, not number of subjects.

TABLE 1

MEAN SCORES ON FOUR TYPES OF SOCIAL ADJUSTMENT IN NURSERY AND NONNURSERY CHILDREN ($N = 42$)

	MEAN SCORE		
Adjustment Scale	*Nursery*	*Control*	t
Activities	5.9	6.7	2.80**
Group	5.7	6.3	2.50*
Authority	6.3	6.5	.93
Personal	5.7	6.5	2.24*

* $p \leq .05$.
** $p \leq .01$.

As can be seen from this table, the control children receive higher mean ratings on each of the scales. The differences between the means are significant for three of the four ratings (activities, group and personal), while the difference on the fourth scale (authority) is consistent in direction but is of insufficient magnitude for statistical significance. These results indicate, then, that, with respect to personal adjustment, participation in group activities, and relations with other children, nonnursery school children are rated by their teachers as better adjusted than those who have attended nursery school. In the case of the adjustment of the child to the teacher, the difference in ratings is not significant.

The Brightness Ratings

Evaluation of the teacher's ratings of the relative "brightness" of the nursery and control children was effected by a nonparametric sign test. Because of the method by which these ratings were obtained, it seemed best to employ a method relatively free from parametric assumptions concerning the variable being measured. Further, since directional differences only were being considered, the sign test for paired replicates seemed fully appropriate.

It was found that in 24 of the 42 paired cases the children who attended nursery school were rated as brighter by their teacher. By sign test this difference is not significant, supporting the conclusion that kindergarten children who have been to nursery school do not differ in *rated* intelligence from those who have not.

Adjustment and Intelligence

To determine the extent to which the various adjustment ratings covaried with the teachers' ratings of the children's brightness, point biserial correlations were calculated between the brightness ratings and each of the four adjustment ratings. For the purpose of this analysis, the children were divided regardless of nursery school attendance into two categories, one consisting of those children the teacher rated as "brighter" and the other including the children she did not so rate. Point biserial *r*s were then computed between the dichotomous ratings and each of the four adjustment scales. The resulting correlations are presented in Table 2.

TABLE 2

POINT BISERIAL CORRELATION BETWEEN INTELLIGENCE AND FOUR ADJUSTMENT RATINGS ($N = 42$)*

Scale	r
Activities	.22
Authority	.13
Group	.27
Personal	.19

* $r \geqq .22$ for significance at $p = .05$.

It can be seen from this table that only two of the four correlations are significantly different from zero and that both of these indicate, at best, only a slight relation between the brightness ratings made by the kindergarten teacher and her ratings on the four adjustment scales. It must be emphasized here that this finding has no relevance to any possible relations between intelligence and adjustment. Any such correlation in the present study would necessarily approach zero as a consequence of the matching procedure which may be presumed to have controlled effectively for any correlation between intelligence and adjustment. Rather, the point of the analysis performed was simply to determine whether children rated "brighter" would exhibit correlative variations in rated adjustment. From the results reported in Table 2 it may be concluded that such seems not to be the case.

As a further check on the relations between the adjustment ratings and the teachers' ratings of brightness, median tests were performed (17). No significant relation was found between brightness and any of the four adjustment ratings, as might be expected from the results of the biserial correlations.

Hence, it would appear that, when actual variations in IQ are controlled, the teachers' ratings of adjustment vary independently of their perceptions of the relative "brightness" of the children rated.

Relations Among the Scales

To determine the degree of relation among the scales, product-moment correlations were computed between the several adjustment ratings. Each type of rating was correlated with each other yielding a matrix of six correlations as reported in Table 3. In calculating each of these correlations, ratings for all 84 children in the sample were pooled regardless of whether they had been to nursery school.

TABLE 3

PRODUCT-MOMENT CORRELATIONS BETWEEN THE FOUR ADJUSTMENT RATINGS ($N = 84$)*

Scale	*Activities*	*Authority*	*Group*
Authority	.54	..	..
Group	.75	.77	..
Personal	.55	.48	.52

* $r \geqq .28$ for significance at $p = .01$.

It may be seen from Table 3 that relations ranging from moderate to strong exist between the teacher's ratings of the various facets of over-all adjustment selected for study. Since the teacher in every instance is rating "adjustment" and is furthermore rating the adjustment of a single child, it would be expected that substantial correlations would exist among these ratings. It may also be inferred from the obtained correlations that kindergarten children tend to exhibit a considerable consistency in the various aspects of their personal and social adjustment, at any rate as these are evaluated here. This finding is, of course, not especially surprising.

Finally, the fact that the correlations reported in Table 3 are rather less than unity (the mean r is approximately .60) is consistent with the assumption made in the construction of the adjustment rating scales, i.e., that each of the scales is tapping an aspect of general adjustment at least partially independent of the others. Therefore, it would appear reasonable to consider the results of the previous analyses in these terms.

Discussion

The results of the present study quite clearly fail to support the hypothesis that nursery school attendance will enhance later school adjustment, at least relative to comparable nonattending children. On the contrary, the indications are that children who have not been to nursery school are perceived by their teachers to be generally better adjusted than are children who have attended such schools.

It warrants emphasis that this study does not demonstrate "absolute" differences in adjustment, excepting insofar as these are reflected in the way the teacher perceives the child. At any rate, it should be clear that other kinds of observers and/or different criteria might yield different results. If one assumes (as in the case in this paper) that adjustment is determined by evaluative operations of observers, it follows that characteristics *both* of the evaluated individual and of the evaluator will be relevant. In other words, assessments of the adjustment of individuals will vary as some function of the characteristics of those individuals in combination with characteristics (including criteria) of the individuals doing the evaluating. As a result, from the same raw data, two observers could arrive at different judgments concerning the "adjustment" of a person. In fact, Tindall (18) has demonstrated that remarkably low correlations exist among different criteria of adjustment.

These points may, in part, account for some of the disparate results reported in the literature. For example, some studies have employed behavior check lists which are completed by the teacher and the results of which are then pooled in some fashion to provide an index of general adjustment. In effect, the investigator provides the adjustment criteria according to his conceptions of the requisites of "good adjustment" and may, therefore, arrive at a judgment different from that which would have been made by the teacher had he or she rated "adjustment" rather than particular characteristics.

As a related point, it might be noted that a judgment of adjustment obtained by simply summing a number of segmental attributes may be quite different from a judgment which takes account of the *gestalt* of such attributes in a given individual. Taken individually, a particular collection of traits may well suggest "good adjustment," but when combined in a given individual it may not be optimal for "good adjustment." Thus, a global rating may yield very different results from a simple sum of attributes.

All of this suggests that a highly significant line of future research pointing toward clarification of some of these issues will address itself to investigation of the factors which enter into the kinds of evaluative operations alluded to above. Until a great deal more information is available, it would seem to be a more meaningful approach to deal directly with the results of these judgments for such purposes as the problem investigated here and to partition the problem into two parts: the evaluative judgments themselves, on the one hand, and the factors which determine them, on the other.

The brightness ratings obtained here are broadly consistent with Page's (16) and other IQ studies. From the failure to find any significant differences it may be concluded that the teacher's perceptions of a child's "brightness" and her perceptions of the child's adjustment are independent dimensions.

While it is not at all simple to account for the results obtained in this investigation, we might note some of the seemingly more plausible hypotheses. Firstly, it is generally true that nursery school offers to the child a freedom and lack of restriction which may run counter to the somewhat greater conformity demanded in the kindergarten. The use of individual initiative seems to be stressed in many nursery schools whereas in kindergarten there are more directed types of activity. Those behavior patterns acquired in nursery school could influence the kindergarten teacher's perception of the child's adjustment.

It is also possible that redundancy of kindergarten and nursery school activities may lead to some primitivation of behavior. A child may become a problem in kindergarten, who has not been one before, because of the repetition of games, songs, naps, art, rhythms, etc. He may show problematic behavior because he is sincerely bored or because this previous knowledge of the curriculum may make him want to show his classmates that he knows all about it and variously try to enhance his status. This interpretation is consistent with Allen and Masling's (1) failure to find significant differences until second grade, differences which then were in favor of nursery school children. A child may take several years to "outgrow" this sort of nursery school-related difficulty, but by the time he reaches second grade the situation may reverse itself and actually become manifest in the form of better adjustment. It would naturally be the brighter children who would be affected most adversely by redundancy, and, since the mean IQ of the present nursery group was 118, it may well be that any positive effects of nursery school would be delayed in demonstrating themselves until there was no longer any repetition in the curriculum.

The rating of adjustment to the teacher showing no difference between nursery and nonnursery children is consistent with these notions. Even though the nursery child may be poorly adjusted for the first few years following nursery school, his relations with the teacher need have nothing to do with his personal adjustment or with his adaptation to kindergarten. He may still manifest friendly relations with the teacher and at the same time be bored with the repetitive activities.

Then too, as a related point, nursery school may foster development of habit patterns, not necessarily bad in the long run, but which may in a *young* child lead to behavior resulting in perception by the teacher of poor adjustment. The child may not "actually" be more poorly adjusted but may strike the teacher that way because of his seeming nonchalance and/or independence. This may be why the teacher rated nursery children lower in group, personal, and activity adjustment. The child may regain his former interest and activity in new things a little later on in school. For example, Ezekiel (7) has noted egocentricity to increase with length of stay in a nursery school. Andrus and Horowitz (3) found that in six out of seven schools the correlation between insecurity and length of time spent in a nursery school was positive as measured by adult raters. Jersild and Mackey (13) note an increase in children's fights and quarrels with nursery school attendance. It may be habits such as these on which the teacher is basing her opinion. Of course, these factors may indicate poor adjustment in something like a clinical sense, but there are other things to be considered.

Apart from the kinds of issues already discussed, it is highly probable that a selective factor influences the enrollments of nursery schools. A study of 100 applicants from the waiting list of a public school nursery school in a midwestern city showed that 51 of the parents believed their children needed supervised play. Eighteen thought the nursery school educational program would be good for the child. Fifteen were employed. Seven said they had an only child and wanted playmates for him. Other reasons were "middle child trouble," "lack of adequate play space," "inability to get along socially," "his older sister enjoyed nursery school," and "we want to broaden our knowledge of child rearing" (15).

These reasons all sound very good, but there probably are others which the parents would not put down on the applications, such as: poor marital situations, parents who feel they are failing the child emotionally, and parents who feel it is a baby-sitting center while they pursue their own interests. A higher proportion of problem children (and/or children from "problem families") probably find their way to nursery school as it is set up at present. Parents encounter difficulty managing such children at home and feel that maybe nursery school is the answer, since one of the functions of nursery school is a counseling service. Some mothers may just want to dispose of the problem child for a few hours a day to settle their own nerves, without any regard for the therapeutic value of a nursery school. Voas (20), in fact, finds more emotional problems and remedial academic difficulties in children who have attended nursery school than in children who have not. It may be that nursery school did not make them problems but that they already were before starting preschool. It is even possible that such children are better adjusted than they would have been had they not attended school.

It is obviously not possible at present to choose among these alternative hypotheses or others which may occur to the reader. At any rate, it is clear that numerous research problems remain to be dealt with in this provocative area.

SUMMARY

This study compared teachers' ratings of social adjustment in 84 kindergarten children equally divided between those who had nursery school training and those who had not. Each nursery child was matched with a nonnursery child from the same kindergarten for sex, IQ, socioeconomic status, and ordinal position in the family. The appropriate teacher was then asked to rate both children separately on their adjustment to group activities, to the teacher, to other children, and on over-all personal adjustment. In addition, a brightness rating was included which asked the teacher to mark which child of the pair she felt to be "brighter." The teacher was kept ignorant of the purpose of the study.

The basic results were as follows:

1. Nonnursery children were perceived by their teacher to be better adjusted than nursery school children in personal adjustment, relations with other children, and participation in group activities. Adjustment of the child to the teacher did not produce significant differences in rating, although the direction was consistent.

2. Children who had attended nursery school did not differ in intelligence, as rated by their teachers, from those who had not.

3. The teacher's ratings of a child's adjustment varied independently of her perceptions of the relative "brightness" of the children rated.

4. Intercorrelations among the four adjustment scales ranged from moderate to high, indicating, on the one hand, a high measure of consistency in these ratings and, on the other hand, a measure of independence.

Various interpretations of these results are offered along with suggestions for further research designed to clarify some questions raised.

REFERENCES

1. Allen, G., & Masling, J. An evaluation of the effects of nursery school training on children in the kindergarten, first and second grades. *J. educ. Res.*, 1957, 51, 285-296.
2. Ammons, R. B., & Ammons, H. S. *Full-Range Picture Vocabulary Test.* Psychological Test Specialists, 1948.
3. Andrus, R., & Horowitz, E. L. The effect of nursery school training: insecurity feelings. *Child Develpm.*, 1938, 9, 169-174.
4. Bonney, M., & Nicholson, E. L. Comparative social adjustments of elementary school pupils with and without preschool training. *Child Develpm.*, 1958, 29, 125-133.
5. Cushing, H. M. A tentative report on the influence of nursery school training upon kindergarten adjustments as reported by kindergarten teachers. *Child Develpm.*, 1934, 5, 304-314.

6. Dixon, W. J., & Massey, F. J. *Introduction to statistical analysis.* McGraw-Hill, 1957.

7. Ezekiel, L. F. Changes in egocentricity of nursery school children. *Child Develpm.*, 1931, 2, 74-75.

8. Fisher, M. L. Measured differences between problem and nonproblem children in a public school system. *J. educ. Sociol.*, 1934, 7, 353-364.

9. Guilford, J. P. *Psychometric methods.* McGraw-Hill, 1954.

10. Hattwick, D. The influence of nursery school attendance upon the behavior and personality of the preschool child. *J. exp. Educ.*, 1936, 5, 180-190.

11. Havighurst, R. J., & Neugarten, B. *Society and education.* Allyn-Bacon, 1957.

12. Hollingshead, A. B. *Two factor index of social position.* Yale Univer. Press, 1957.

13. Jersild, A., & Fite, M. Children's social adjustments in nursery school. *J. exp. Educ.*, 1937, 6, 161-166.

14. Kawin, E., & Hoefer, C. *A comparative study of a nursery school versus a nonnursery school.* Univer. of Chicago Press, 1931.

15. Moustakas, C., & Berson, M. P. *The nursery school child care center.* Whiteside, 1955.

16. Page, J. D. Effect of nursery school attendance upon subsequent IQ. *J. Psychol.*, 1946, 10, 221-230.

17. Siegel, S. *Nonparametric statistics for the behavioral sciences.* McGraw-Hill, 1956.

18. Tindall, R. H. Relationships among measures of adjustment. In L. Gorlow & W. Katkovsky (Eds.), *Reading in the psychology of adjustment.* McGraw-Hill, 1959. Pp. 107-115

19. VanAlstyne, D., & Hattwick, L. A. A follow up study of nursery school children. *Child Develpm.*, 1939, 10, 43-72.

20. Voas, W. H. Does attendance at the Winnetka Nursery tend to raise the IQ? *Yearb, nat. Soc. Stud. Educ.*, 1940, 39(II), 363-377.

21. Walther, J. The influence of nursery school education upon behavior maturity. *J. exp. Educ.*, 1939, 8, 164-165.

PREDICTION OF SCHOOLAGE INTELLIGENCE FROM INFANT TESTS[1]

Sibylle K. Escalona* *and* Alice Moriarty

Albert Einstein College of Medicine

When applied at age levels below 18 months, the term "intelligence test" is misleading. At the preverbal stages of development the group of functions designated as intelligence, and measured by intelligence tests applicable at schoolage and beyond, has not yet emerged. The true relationship between that which is measured by infant tests and that which we later call intelligence remains largely unknown. What little we do know, however, suggests that infant IQs or DQs do not reflect functions in direct continuum with the group of functions later assessed by intelligence tests. Or, as Bayley put it in a fairly recent review of the problem, "It is now well established that we cannot predict later intelligence from the scores on tests made in infancy" (2, p. 806). With one important exception to be mentioned below, no significant correlations have been reported between test scores obtained at less than 1 year of age and intelligence test scores obtained in middle childhood and beyond.

Curiously, these consistently negative results have had an impact upon social practice (especially in the field of adoptive placement), but they have not diminished scientific concern with the area of infant testing. On the contrary, the most recent decade brought a revival of interest, as attested by a larger number of publications and by the increased time allotted to the topic at many professional and scientific meetings. In large measure the

---

* Department of Psychiatry, Albert Einstein College of Medicine, Yeshiva University, Eastchester Road and Morris Park Avenue, New York 61, New York.

[1] Originally presented as a paper at Biennial Meeting, Society for Research in Child Development, Pennsylvania State University, March, 1961.

Child Develpm., 1961, 32, 597-605.

current interest in infant testing is an outgrowth of the general interest in very early phases of human development. However, in addition many clinicians experienced in infant testing and knowledgeable about development as a whole continue to believe that infant tests elicit behavior closely related to central developmental processes. Therefore, they also consider that differences in the test performance of infants must refer to characteristics of the organism which ought to affect learning potentials and hence later intelligence test scores. And indeed for Gesell DQs obtained at 40 weeks, Knobloch, Pasamanick, and their associates have consistently reported significant correlations with Gesell scores at 3 years and with Stanford-Binet scores at that and somewhat later ages (4, 5). Two conclusions have been agreed upon by all investigators: (a) Infant scores obtained during the first half year of life do not predict later intelligence. (b) Infant tests are relatively most effective in discriminating between those who will later show neurological and intellectual deficit and those who will not, whereas they are less effective in predicting intelligence level within the normal and superior ranges.

We count ourselves among those investigators who consider that infant tests are the best available method for the systematic assessment of developmental status and for the measurement of mental growth. Our special interest is focused on very early developmental processes and their relationship to more stable patterns of mental functioning at later ages. We were therefore motivated toward further study of the possible predictive significance of infant test scores in precisely the age range where correlations with later IQs are known to be insignificant. Similarly we wished to learn more about the relationship between individual differences in test scores manifested during the first half year of life and within the normal range and comparable differences in IQ scores obtained at an age when the IQ has become as relatively stable as it ever will be.

Before turning to the empirical findings, it will be useful to turn attention away from the usual criteria of test validation and test reliability (all of which presuppose either the measurement of existing functions or the measurement of a straightforward antecedent–consequent relationship). Instead, we would call attention to currently accepted theory about basic processes of mental development.

Most of us do not view intelligence as a fixed potential or a trait, latently present in early infancy and gradually unfolding with maturation. Certainly the authors are not alone with their assumption that intelligence must be viewed as an interaction phenomenon. With apologies to Piaget's and other developmental theories we shall crudely describe intelligence as the result of a continuous stream of transactions between the organism and the surrounding field, which field is the sum total of physical and social environmental conditions. Since the organism brings to each adaptive act certain properties, both those of an intrinsic biological nature and those which are the result of the impact of previous experience upon the organism, it is correct to say that intelligence development is at all times dependent upon what the organism is like. Yet, since it requires environmental circumstance to mobilize the organism, and since the kind of transaction which develops depends on the objective content of that to which the organism must adapt, it is equally true to say that the development of intelligence depends at all times on the experience encountered by the growing child. If Gordon Allen's use of the term "genotype" is modified so as to include properties of the organism that may have been acquired pre- or paranatally, our view approximates his in emphasizing that the phenotypic expression of the same genotypical characteristic will vary enormously, depending upon the matrix of developmentally relevant outer circumstance (1).

If all of this is reasonable, it follows that a manifest deficit in sensorimotor functioning should enter each adaptive transaction and should limit or distort the development of intelligence. It would further follow that exceptionally good organismic equipment, as manifested by acceleration or superiority of sensorimotor function, should equally enter all transactions and enable the organism to derive maximal benefit from available stimulatation and thus show superior intelligence later. Such a view is compatible with the expectation that average and superior endowment may be suppressed by limiting or actively traumatic life experience—as indeed we know can be the case. It is also compatible with the expectation that unusually facilitating experience can enhance the emerging intelligence at any level of initial endowment. But our view also requires that such facilitation occurs within biologically determined limits. No amount of facilitating experience can be expected to produce a level of intelligence functioning that transcends the limits set by biological givens. In the absence of unusual limitation and of very unusual facilitation, under what Hartmann (3) termed the "average expectable environment," marked differences in the level of sensorimotor functioning should then correspond to marked differences in later intelligence functioning.

The group of children on whom we obtained infant test scores some 10 years ago, and who were retested at early schoolage, come close to meeting this last requirement. Our subjects were originally selected for "normalcy" by medical, social, and developmental criteria.[2] They have grown up in a midwestern town which provides a relatively benign and consistent milieu.

Of the original sample of 128 infants so selected, 70 were subsequently tested by my collaborator and by one other psychologist in connection with two separate research projects.[3] In investigating the degree to which infant test findings bear a relationship to later intelligence, we were anxious to use only test scores obtained under optimal circumstances. This led to considerable shrinkage in the sample, as we discarded all instances where test reports expressed doubts of the validity of test results. Thus, we shall report on a population of 58 children who were first tested between the ages of 3 weeks and 33 weeks and tested again between the ages of 6 years, 2 months, and 9 years, 3 months. The infant tests were distributed fairly regularly between the ages of 4 weeks and 32 weeks, but the great majority of retests used were administered between 7 and 8½ years. Both the Cattell Infant Intelligence Scale and the Gesell Developmental Schedule were given to each infant subject. In addition, for each infant, we had available an evaluation of developmental status based on both test protocols jointly. This "clinical appraisal" took into account qualitative factors as well (such as which areas of functioning were relatively advanced or lagging and "good failures" or minimal successes) and was a part of the *original written protocol*. It proved that these clinical appraisals could be treated as ratings, distinguishing between low average, average, high average, and superior, as judged by the total infant performance. For the schoolage assessment we used only the Wechsler Intelligence Scale for Children (WISC), although for about half the subjects Stanford-Binets and other tests are also available.

---

[2] The sample is selected so as to include children who met criteria of normalcy; it is not necessarily a representative sample.

[3] We are indebted to Dr. Herman A. Witkin, Professor of Psychology of State University of New York, for permission to use intelligence test results obtained as a part of his research on the origins of different perceptual styles.

Figure 1 describes the sample in terms of intelligence ranges, as determined by each of the three tests used. Note that in order to make meaningful discriminations in this sample, we departed from convention and called "low average" the range from 90 to 99 IQ or DQ points. At the time of initial testing two DQ scores between 86 and 90 were obtained, but no IQ scores below 90 were obtained at either time. The remarkable distribution of Cattell scores (58 per cent superior) reflects a statistical attribute of this instrument, which assigns a weight of .2 to each of five items so that a 4-weeks-old who passes all items at the adjacent level obtains an IQ of 200. The Gesell DQ scores come closer to the distribution as defined by WISC IQs.

In line with what was said earlier, our expectation of infant tests was *not* that they should predict later IQs. Instead, we wished to learn what portion of the variance they may contribute to later IQ scores. Or, more accurately, we should like to learn whether assessment of developmental status in infancy significantly defines the *range* of intelligence that may be expected to emerge.

From this vantage point the most common techniques of assessing the predictive effectiveness of tests become inappropriate. If nonetheless we report a few of the parametric measures, it is for the sake of comparison with other studies. As expected, the product-moment correlation between Cattell IQs and WISC IQs is totally insignificant (—.045), and the same is true for the correlation between DQ scores and later IQs (.077).

As will be shown later, it proved that in our sample a relationship between infant and schoolage scores could be demonstrated only for subjects whose infant scores were obtained at ages 20 weeks and above. That is, no matter how one modifies one's expectations, those infant tests given to

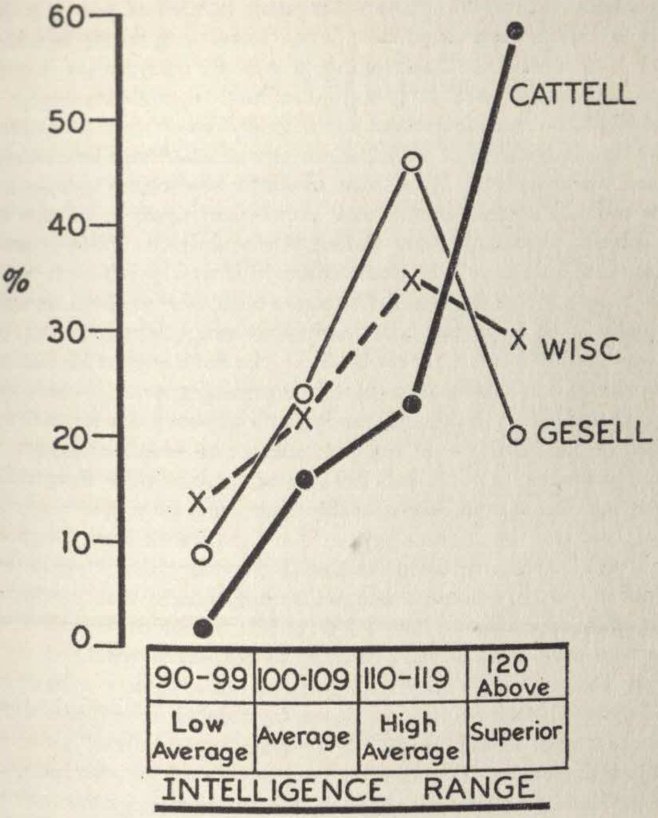

FIGURE 1—Distribution of sample over intelligence ranges in percentages as assessed by Cattell IQ, Gesell DQ, and WISC IQ.

normally progressing babies between the ages of 3 and 17 weeks bore a purely coincidental relationship to WISC IQs (except, of course, for the fact that not one of these 27 subjects obtained a WISC below 91). We therefore obtained correlations for the 31 subjects who were originally tested at 20 weeks or above and found that for the Cattell IQs $r = .212$, whereas for the Gesell DQ $r = .253$. Neither of these values is of interest. One other among the conventional measures may be mentioned, again primarily for comparison with other published values. A Spearman rank-order correlation using Gesell DQs and WISC IQs proved completely insignificant for the total population ($r_s = .062$), and very nearly so for the 31 subjects tested between 20 and 33 weeks ($r_s = .20$, which is at the .10 level of significance).

In these terms then, our group of children behaved very similarly to what has been reported for many other small samples of normal children. In order to test our expectation that infant tests do *something* to define the limits of later intelligence functioning, it was necessary to use a measure which would rank subjects in groups according to intelligence range. The Kruskal-Wallis one-way analysis of variance by ranks proved a suitable tool. It allows the classification of subjects into grouped rankings based on infant scores and works with the WISC raw scores for the second measure.

This test was applied to the total population in three different ways. When subjects were ranked by Cattell IQs against the WISC scores, no significant association could be demonstrated ($H = 1.73$, $df = 2$, $N = 56$,[4] $.50 > p > .30$). When the Gesell DQ scores were used to rank our subjects, they equally failed to predict later intelligence range when the total population was used ($H = 1.3$, $df = 3$, $N = 58$, $.80 > p > .70$). Lastly, the total population was ranked in terms of the clinical appraisal which, as will be remembered, was a judgment made at the time of the original testing and based on the infant's total test performance on both instruments, with attention to qualitative features.[5] For the total population it proved that a modest but demonstrable relationship does exist between these clinical appraisals and later intelligence range. The significance level (of an $H = 4.11$ with $df = 2$) was between .10 and .20, actually very close to the .10 level. Such as it is, the statistic is somewhat misleading in that the association between infant appraisal and later IQ is contributed almost entirely by those subjects who were tested at the age of 20 weeks and above.

On the basis of our theoretical orientation and of clinical experience we believed that—with the exception of neurologic deficit which here does not apply—infant tests would have little bearing on later intelligence when limited to a single test during the first four months of life. We had guessed that 16 weeks might be the cut-off point, for it is at this age that such more cortically determined functions as voluntary movement emerge with clarity. It proved, however, that for this sample the critical age was 20 weeks. A great many different calculations (which were hardly necessary since simple inspection left little doubt) proved that, for the 27 subjects between 3 and 17 weeks at the time of first testing, chance alone accounts for the later distribution of WISC IQs.

The Kruskal-Wallis rank-order procedure was applied to the small sample of children 20 weeks and above. When the Cattell scores were used, it was now possible to demonstrate a modest relationship as the p value fell between the .20 and .10 level ($H = 3.91$, $df = 2$, $N = 30$). We expected the Gesell Schedules to be the better instrument, chiefly because it taps a larger variety of relevant sensorimotor functions. On the other hand we consider the DQ an unfortunate numerical expression as it bears an ambiguous relationship to an irregular number of items at each age level. It is perhaps in keeping with these different attributes of the same test that, as reflected by the Kruskal-Wallis, DQ rankings showed a slightly closer relationship to later WISC scores than did Cattell IQs in that the p value fell at exactly the .10 level ($H = 6.27$, $df = 3$, $N = 31$).

---

[4] Only two subjects fell into the low average group by the Cattell criterion; these were omitted from the calculation.

[5] These clinical appraisals do not refer to those cases where it was judged that test performance was for some reason unrepresentative of the child's developmental status. All such instances were excluded to begin with and were the major reason for the reduction in sample size from 70 to 58.

When clinical appraisals were ranked against later WISC scores, and the sample was limited to subjects who were no younger than 20 weeks of age, a significant degree of correspondence emerged. The p value for this calculation falls between the .01 and .02 level of significance ($H = 8.25$, $df = 2$, $N = 31$). The sample is too small to permit of generalization, and the Kruskal-Wallis technique, it should be remembered, does no more than to demonstrate the significance of the difference between different categories—in this case intelligence ranges. Within these limitations we find food for thought in the observation that a qualitative evaluation of only test performance, i.e., not of situational factors that may have influenced performance, is considerably more effective in predicting intelligence *level* than are test scores.

It occurred to us that the distinctions between low average, average, high average, and superior may still be finer than the discriminating powers of infant tests with respect to later IQ scores. We therefore reassigned our subjects into only two categories, average and superior. All IQs or DQs in the range from 90 to 114 were now called average, those 115 and above were called superior. With two exceptions (which were dropped from the sample) the clinical appraisals were so worded as to make it possible to distinguish those infants who had been judged as possessing markedly superior endowment from those who showed small degrees of developmental superiority. Working only with babies tested at 20 weeks and above, we wished to compare the proportion of subjects who, from infancy to school-age, remained in the same category with those who had shifted categories. With an N reduced to 29, such a comparison offers statistical embarrassment. Short of an analysis of variance (which is not applicable to the clinical appraisals) there is no way of including the actual distance between WISC scores and the predicted range in the calculation. Thus, an IQ change from 114 to 115 constitutes a shift from average to superior just as would be the case with a shift from 100 to 130. It is perhaps for this reason that the Gesell scores, when grouped only into average and superior, failed to predict the subject's placement into corresponding categories by the WISC criterion. However, when clinical appraisals were used the same simple comparison of those who stayed within the range established in infancy with those who moved into the other range proved highly significant. Twenty-three or 79 per cent of the 29 subjects remained in the same category. A binomial test of proportion applied to this result yields an error estimate at the .002 level.

It may be of incidental interest that in this small sample all of the six shifts occurred among the 22 subjects originally assessed as superior. Our clinical appraisals tended to err in overestimating endowment. They did not, in this sample, underestimate the range of later intelligence functioning.[6] (Yet for the superior group alone the proportion who remained within category was significant at the .034 level).

SUMMARY AND CONCLUSIONS

Fifty-eight children selected for normalcy by medical, social, and developmental criteria were tested between the ages of 3 and 33 weeks. For each infant we obtained (a) Cattell IQ scores, (b) Gesell DQ scores, and (c) a "clinical appraisal" based on total test performance. These subjects were again tested between the ages of 6 years, 2 months, and 9 years, 3 months, at that time receiving the Wechsler Intelligence Scale for Children (WISC).

On theoretical grounds, and in view of previous findings, parametric measures of correspondence between infant scores and later intelligence test scores were not expected to yield significant values. In this respect, results confirmed our anticipation. The hypothesis under investigation was that, in the absence of major environmental variation, infant tests will predict the *general level* of later intelligence functioning within the average and superior ranges. Further, we compared the relative predictive power in terms of intelligence ranges of Cattell IQs, Gesell DQs, and a clinical appraisal of both test protocols in combination.

It proved that, when infant tests were administered to subjects younger than 20 weeks of age, no method of appraisal predicted later intelligence range. For subjects tested between 20 and 32 weeks of age inclusive, both the Cattell IQ and the Gesell DQ bore a positive relationship to later intelligence level, as assessed through a one-way analysis of variance by ranks. However, this relationship exceeds chance at the .10 level and is therefore of limited predictive value. When the same test is applied to subjects tested at 20 weeks and above but ranked on the basis of clinical appraisal, infant assessment significantly predicted differences in later intelligence range ($.02 > p > .01$). When infant assessments were examined for their ability to distinguish merely between subjects who would later be of average or of above average intelligence, clinical appraisal (but neither of the test scores) achieved this discrimination at a highly significant level (.002).

We concluded that infant tests administered under optimal conditions at 20 weeks of age and above can approximate the prediction of later intelligence range if environmental conditions are relatively stable and benign. It was further concluded that, for children in this age range, a clinical appraisal based on total test performance on two test instruments is a better predictor of later intelligence range than are scores obtained from either the Cattell or the Gesell tests alone.

---

[6] This direction of change is in accord with results reported by Knobloch *et al.* (4, 5) and can be thought of in terms of their hypothesis to the effect that detrimental environmental influences reduce intellectual achievement of which the organism is potentially capable.

It is suggested that research dealing with the contribution of biological and maturational factors to the development of later intelligence may become more rewarding once it is assumed that developmental examinations are not tests of intelligence. If infant tests are viewed as assessing only one among several determinants of intelligence, their usefulness for research purposes is not determined by the usual criteria of test validation, and the failure to predict later IQ scores does not relegate infant testing to the limbo of useless enterprises.

REFERENCES

1. ALLEN, G. Intellectual potential and heredity. *Science*, 1961, 133, 378-379.
2. BAYLEY, N. On the growth of intelligence. *Amer. Psychologist*, 1955, 10, 805-818.
3. HARTMANN, H. *Ego psychology and the problem of adaptation*. International Universities Press, 1958.
4. KNOBLOCH, H., & PASAMANICK, B. The distribution of intellectual potential in an infant population. In *The epidemiology of mental disorder*. American Association for the Advancement of Science, Publication No. 60, 1959. Pp. 249-272.
5. KNOBLOCH, H., & PASAMANICK, B. An evaluation of the consistency and predictive value of the 40 week Gesell Developmental Schedule. In C. Shagass & B. Pasamanick (Eds.), *Child development and child psychiatry*. Psychiatric Research Reports of the American Psychiatric Association, 1960, No. 13. Pp. 10-31.

COMMENT ON "THE RELATIONSHIP BETWEEN RIGIDITY-FLEXIBILITY IN CHILDREN AND THEIR PARENTS"

JOHN WILLIAM ASHER*

University of Pittsburgh

In the 1959 article by Blum (1) intercorrelations among mothers, fathers, and children on measures of rigidity-flexibility were presented. In addition to the three simple two-variable correlations among the three groups, multiple correlation of mothers' and fathers' rigidity with their child's score was obtained (*see* Table 2 of the cited article).

There were no significant correlations among the three variables taken two at a time. A multiple correlation of .459 was found for mothers' and fathers' scores taken as the dependent variables and the children's scores taken as the independent variable with 17 scores on each variable.

The author concludes, "These findings support the prediction, namely, that in a normal sample the child's relative rigidity-flexibility pattern is positively related to this variable in his parents. Although no significant relationships were found for the child and either one of his parents separately, a significant relationship was found when both parents were considered. Such findings should further support investigations which attempt to understand and measure the influence and effect of *both* parents, as a family entity, upon the growth and development of the child" (1, pp. 303-304).

To this reader, the significance levels presented with the size of the multiple correlation obtained, especially with the relatively small size of the sample, seemed unusual. A standard statistical table was consulted, Table D, Coefficients of Correlation and *t* Ratios Significant at the 5 Per Cent Level . . . and at the 1 Per Cent Level . . . for Varying Degrees of

---

* School of Education, University of Pittsburgh, Pittsburgh 13, Pennsylvania.

Child Develpm., 1961, 32, 607-608.

Freedom (2, p. 609). For a three-variable relationship with 17 scores on each, 14 degrees of freedom remain, $17 - 3 = 14$. Thus, the table is entered with 14 *df* and with three variables. We find that a multiple correlation of at least .590 is needed to reject the null hypothesis of no correlation at the 5 per cent level and .694 at the 1 per cent level. Obviously, the multiple correlation reported is not of this size. Even a simple correlation of this size would not be considered significantly different from zero with just 17 observations.

Thus, the report that ". . . a significant relationship [between child and parent's rigidity-flexibility measures] was found when both parents were considered" (1, pp. 303-304) is in error. From the data provided it must be concluded that there is *no* relationship. The further theorization derived from these data would also need to be reconsidered.

However, it is believed that the opposite conclusions from those reported should not be accepted. It would appear that the small number of observations would almost invariably lead to indeterminant results unless significant findings ensue. The results must be accepted for the moment. However, for experimenters, the viewpoint should be that an interesting area of research has just started to be investigated and needs further research with larger numbers of observations before more definitive conclusions can be reached.

These comments do not of course detract from the other findings and developments of the study.

References

1. Blum, A. The relationship between rigidity-flexibility in children and their parents. *Child Develpm.*, 1959, 30, 297-304.
2. Guilford, J. P. *Fundamental statistics in psychology and education.* McGraw-Hill, 1950.

FAMILY ROLE CONCEPTS OF CHILDREN AGES SIX TO TEN[1,2]

WALTER EMMERICH*
University of Colorado

A number of social scientists have emphasized the importance of role perception as a determinant of social action (1, 9, 15, 18, 20). It is believed that role discriminations help define the constancies anticipated in each person's behavior during the ensuing interaction. Insofar as such anticipations function to mediate and guide purposeful activity, a thorough understanding of their nature is essential for predicting behavior. Very little is known about the dimensions children use to define dyadic relationships within the family. In an earlier study by the writer (7), this problem was investigated in children of 4½ to 6 years. The present study extends the investigation to an older sample, with the purpose of describing intragroup role consensus and differences in consensus among subgroups classified according to sex and age.

The descriptive aim of this study is pertinent to several areas of developmental social psychology. First, any subgroup regularities in role perception will in part describe the typical child's social space and thereby become parameters in predicting his behavior. Second, the functioning of a social system is presumably dependent upon the extent of agreement on role definitions among its members. By comparing normative data on children and parents, it should be possible to determine characteristic areas of definitional agreement and disagreement between parent and child and to relate these to indices of family integration. Third, sex and age variations in role consensus have bearing on theories of developmental change in the cognition

---

* Department of Child Development and Family Life, Purdue University, Lafayette, Indiana.

[1] This investigation was supported by research grants M-2853 (A) and M-3844 (A) from the National Institute of Mental Health, United States Public Health Service.

[2] The writer wishes to express his appreciation to Marilyn McRae and Ronald Cunat for their assistance in this study and to Assistant Superintendent Melvin Wiesley, Principal Keith Dillman, and the teachers of the Martin Park Elementary School, Fairview District, Boulder, Colorado, whose cooperation made this study possible.

Child Develpm., 1961, 32, 609-624.

of social structures, such as that of Parsons (18). Fourth, if role conceptions are influenced by cultural arrangements which vary with such factors as social status, sex, and age, it becomes necessary to have subgroup norms in order to evaluate properly the degree of "normalcy" of a particular individual's role perceptions. Finally, as Parsons (18) and Brim (4) point out, knowledge about specific intrafamilial role relationships is essential for understanding the precise nature of more complex role processes. To illustrate this point, consider the commonly held belief that sex-identity is ordinarily acquired by the child's taking one of his parents as a model. Assuming that this belief is correct,[3] the problem still remains of specifying which role relationships serve as referents of the modeling process. For example, in the case of the girl who takes her mother as a model in play, to what extent does her play behavior reflect the mother's relationship with the girl, the father, a sibling, a person outside the family, or some combination of these? Suppose that the girl discriminates age roles, but not sex roles. Then she might take the mother-girl relationship as a model, but she cannot do so for the mother-father relationship because she does not discriminate sex roles. The assessment of children's role concepts, then, is believed to be prerequisite to the analysis of identification patte's fuacquired through a process of role modeling.[4]

It is important to select for study those role dimensions which are likely to be salient for both sexes over an extended age period. The literature indicates that *relative power* and *attitudinal direction* are two dimensions which meet this criterion (3, 8, 13, 14). Relative power is defined here as the amount of implied or explicit control each person has over the outcome of an interaction. Attitudinal direction is defined as the extent of implied or explicit agreement (positive) or disagreement (negative) with respect to the means and/or goals of an interaction. Other definitions used in the study are as follows. A *role* is defined as the behaviors (performances) characteristic of the incumbent of one position toward the incumbent of another position in a social structure. This definition is similar to Gross, Mason, and McEachern's (9) concept of "role behavior sector." A *role concept* is defined as the discrimination of a particular position on a specific behavioral dimension. The referent of a *person concept* is distinguished from that of a *role concept*. A person concept is derived from a particular person's performances of a variety of roles, whereas a role concept is derived from the performance of a typical person or many persons in a single role. There are at least two criteria for operationally distinguishing person and role concepts. For a role, but not for a person concept, there is: (a) generalization to the population of persons who hold the same position; (b) intragroup consensus on the behavior characteristic of a position, when each group member judges a different occupant of the position. This study relies primarily upon the second of these criteria.

---

[3] Bronfenbrenner (5) points out that there is little direct support for this assumption.

[4] This discussion assumes that the child's role concepts and models are assessed from the child's reports, rather than from observations of overt parent and child behaviors. Nevertheless, the point made here would also apply to the latter case, since family roles must in fact differ if differential modeling is to occur.

METHOD

Subjects

Subjects were 225 elementary school children from middle-class families unbroken by separation, death, or divorce. None was adopted. Age groups were formed in such a way as to create nonoverlapping age ranges, approxi-

TABLE I

DESCRIPTION OF SUBJECTS

		AGE (IN MONTHS)		
	N	*Mean*	*Range*	*SD*
6 Year Group				
Girls	21	71.5	66–75	2.7
Boys	17	70.4	64–75	3.4
Total	38	71.0	64–75	3.1
6½ Year Group				
Girls	18	78.9	76–82	2.0
Boys	19	79.7	76–82	1.8
Total	37	79.3	76–82	1.9
7 Year Group				
Girls	19	87.9	83–91	2.8
Boys	20	86.8	83–91	2.9
Total	39	87.3	83–91	2.8
8 Year Group				
Girls	17	96.2	92–104	3.7
Boys	21	96.4	92–103	3.2
Total	38	96.3	92–104	3.7
9 Year Group				
Girls	20	108.7	105–116	3.7
Boys	16	108.7	104–115	3.5
Total	36	108.7	104–116	3.5
10 Year Group				
Girls	16	121.1	117–131	3.7
Boys	21	121.6	117–129	3.3
Total	37	121.4	117–131	3.4
Total				
Girls	111	93.2	66–131	17.2
Boys	114	94.4	64–129	17.6
Total	225	93.8	64–131	17.4

mately the same mean age for both sexes within each age level, and approximately equal *N*s in all subgroups. Six age groups resulted, henceforth called the 6, 6½, 7, 8, 9, and 10 year groups. Subgroup sizes and age statistics are given in Table 1. Approximately 79 per cent of the sample came from socioeconomic levels 1 or 2 in Warner *et al.*'s (occupational) classification scheme (25), and approximately 76 per cent of the sample had either one or two siblings.[5]

Materials

Stylized figure drawings, described in detail elsewhere (7), were prepared on 6¼ by 3¼ in. Hunter Card Master cards in the following pairs: mother-girl, father-boy, mother-father, and girl-boy. In order to mix the presentation of pairs and items, 16 cards were made for the mother-girl and father-boy pairs, and eight cards were made for the mother-father and girl-boy pairs. Each member of a pair appeared on the left and right sides of the cards half the time. The 48 cards were presented to subjects by means of the Hunter Card Master, Model 340.

Procedure

Subjects were tested individually. *S* sat in front of the exposure window of the Card Master, with *E* seated beside him. *E* pointed to the exposed boy-father card and said, "Do you see these people here? Let's play a game with them. Okay? Who says, 'I'm the boy'?" After *S* pointed to the correct figure, *E* said, "Yes, that's a boy just like you" if *S* was a boy, or "Yes, that's a boy just your age," if *S* was a girl. *E* then asked, "Who says, 'I'm the father'?" After *S* pointed to the correct figure, *E* exposed the next card which was the girl-mother pair. Experimenter then asked, "Who says, 'I'm the girl'?" After *S* pointed to the correct figure, *E* said, "Yes, that's a girl just like you," if *S* was a girl, or, "Yes, that's a girl just your age," if *S* was a boy. *E* then asked, "Who says, 'I'm the mother'?" In order to give some initial practice with behavior descriptions similar to the actual test items, the girl-boy card was presented next, with the item, "I want it." The mother-girl card was then presented with the item, "Stop doing that." The test proper then began and consisted of presenting the 48 pairs in a prearranged order, designed to control for a variety of possible response sets. Either pointing to a figure or naming it was an acceptable response. *S* was given as much time as he wanted, and *E* repeated the item whenever *S* did not respond after several seconds or responded ambiguously. Each child was told that he did very well at the completion of testing.

In order to have partial control over the influences of examiner characteristics on the children's responses, a male and a female tested approximately half the subjects in each of the 12 subgroups.

---

[5] These figures are based upon the 157 cases for which exact data were available on socioeconomic level and number of siblings. The remaining 68 subjects were very similar to this group in both respects.

TABLE 2

CLASSIFICATION OF ITEMS ACCORDING TO RELATIVE POWER AND ATTITUDINAL DIRECTION

	High Power	*Low Power*
Positive Attitude		
	1. You can have it.	1. I want it.
	2. That's nice. You did what I asked.	2. I'll do what you say.
	3. You made that very well.	3. I made it very well.
	4. You did what was right.	4. I'll do what is right.
Negative Attitude		
	1. You can't have it.	1. Give me what *I* want.
	2. You'd better do as I say.	2. No. I won't do it.
	3. You didn't make it very well.	3. I can't make it very well.
	4. You did something that is wrong.	4. I'll do something that is wrong.

Measures

The items were similar to those used in the previous study (7) and are classified in Table 2 according to relative power and attitudinal direction. Items having corresponding numbers across columns and rows are role reciprocals on the power and attitude dimensions, respectively. All items were used to assess age-role discriminations of the mother-girl and father-boy pairs. Because of limited time for testing and because of the tendency for younger subjects to assign high and low power items to the adult and child generations, respectively (7), high power items only were used to assess sex-role discriminations of the mother-father pair, while low power items only were used to assess sex-role discriminations of the girl-boy pair. For the mother-girl and father-boy comparisons, the *power discrimination measure* was the total number of assignments of high power items to adult figures and low power items to child figures. For the mother-father comparison, this measure was the number of times the father was chosen on high power items, and, for the girl-boy comparison, this measure was the number of times the girl was chosen on low power items. For the mother-girl and father-boy comparisons, the *attitude discrimination measure* was the total number of assignments of positive items to adult figures and negative items to child figures. For the mother-father and girl-boy comparisons, this measure was the total number of assignments of positive items to female figures and negative items to male figures.

The above measures were based upon the assumption that all items contributed to both dimensions according to the scheme presented in Table 2. A number of scale characteristics, such as small variances and skewed distributions, made it impossible to estimate scale homogeneities and the discrimination power of items with precision. However, crude analyses of part-whole association were performed, using the extremes of those subgroups having maximal variability on the scales. All items were found to contribute to both scales applied to each of the four paired comparisons. The chi square probabilities associated with the 32 phi coefficients ranged from .20 to .001, with a median of .01.

RESULTS AND DISCUSSION

Nonparametric statistics were used in all analyses, and a finding is interpreted if it was significant beyond the .05 level. For analyses of intragroup consensus, chi square was used to determine if more than half the *Ss* fell either above or below the scale point representing equal assignment of both ends of a dimension to each person in a pair. This point on the scale shall henceforth be referred to as "chance." The median test or its extension to several independent variables in a single analysis (26) was applied to differences among subgroups. The sign test was used in the analyses of differential item sensitivities. The index of association was *C* (significance levels only are reported), with distributions divided at or near the median.

Discrimination of Age Roles by Relative Power

There was a marked tendency for the whole sample to use the power dimension to discriminate the mother-girl and father-boy pairs. Scores for these two pairs combined could vary from 0 to 32, with 16 representing chance. Ninety-nine per cent of the whole sample had scores above 16, and the median was 29.5. It is clear that the power dimension is used by children of this age to discriminate age roles, as is the case for younger children (7). This result is highly consistent with Barker and Wright's (3) survey of actual parent-child interactions, in which it was found that dominance and nurturance are the most frequent parent behaviors, whereas appeal, submission, and resistance are the most frequent child behaviors.

Girls discriminated age roles on the power dimension somewhat more than boys ($p < .05$), and this difference did not vary with age. Although this finding is consistent with those suggesting that girls are more sensitive to age grading than boys (6, 10, 21) or that boys are more successful than girls in defying parental commands and restrictions (19), it does not very strongly support either of these interpretations, since boys clearly discriminated age roles by power, and the difference between the sexes was only about one point on the scale. Further, this sex difference did not appear in younger children (7).

Age changes in age-role discrimination by power are presented in Figure 1. The tendency for scores to increase with age is significant for the father-boy ($p < .001$), but not for the mother-girl comparison. There was no interaction between sex and age of child for either comparison. These find-

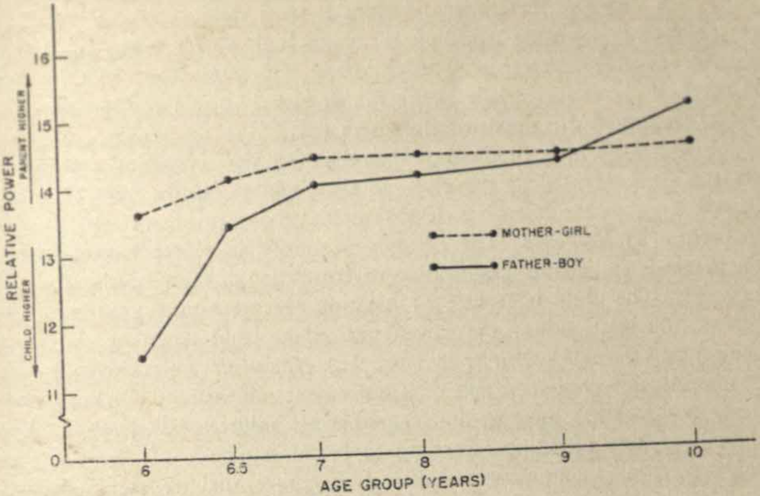

FIGURE 1—Discrimination of age roles by power.

ings are subject to several possible interpretations. If it is assumed that the age trend for the (nonassessed) father-girl comparison would have been similar to that found for the father-boy comparison, then the results signify that children age-type maternal relationships earlier than paternal relationships. Several factors could account for this possibility. Since mothers interact more frequently with children of this age than fathers (3), there is more opportunity for children to learn appropriate age roles in their maternal relationships. However, there is little evidence indicating that the frequency of father-child interactions increases with age during this period (3, 23), as did the father-boy discrimination. Or, if the mother is (or is perceived as) more nurturant or powerful than the father in socializing the younger child, then the child may be more motivated to overlearn appropriate age roles in the maternal relationship. Again, however, although some studies support this interpretation (3, 4, 19), others do not (4, 8, 13, 14). Another possibility is that the child's awareness of the father's high power in other roles increases with age and generalizes to the father-child relationship. This interpretation gains support from the finding, discussed in more detail later, that amount of power attributed to the father vis-à-vis the mother also increases rapidly between the sixth and eighth year (*see* Figure 2). On the other hand, if it is assumed that the age trend for the (nonassessed) mother-boy comparison would have been similar to that found for the father-boy comparison, then the findings signify that children of both sexes perceive that girls age-type earlier than boys. A number of studies do suggest that girls perceive, accept, or are expected to learn age-appropriate child and adult roles earlier than boys (6, 10, 16, 19, 21). Further study is needed to determine which of these interpretations is most valid.

There was a greater tendency to assign high power items to parents than to assign low power items to children ($p < .001$), and this difference showed no consistent variation with sex or age. This is a replication of an earlier finding and is discussed elsewhere (7).

Parsons (18) has hypothesized that there are developmental changes in the specific cues used by the child to discriminate family roles. His discussion suggests the following developmental sequence of age-role discriminations: (a) *Response:* The child is primarily an object having needs; the parent is an agent of gratification or frustration of these needs. (b) *Acceptance:* The child is capable of meeting specific norms governing the parent-child relationship; the parent praises the child when he conforms, and controls the child when he deviates. (c) *Approval:* The child is capable of meeting universal standards of achievement; the parent responds positively to the child's successes and criticizes his failures. (d) *Esteem:* The child is capable of acting according to broad standards of conduct; the parent recognizes appropriate behavior and reproves inappropriate behavior.

The four kinds of items in each cell of Table 2 were formulated to correspond to each of these presumed developmental levels. It was hypothesized that age-role discrimination by power is a function of the interaction between item type and age of child; i.e., lower level items would be age-typed more by younger children, whereas higher level items would be age-typed more by older children. The results did not support this hypothesis, as age-role discrimination by power increased with age at about the same rate for all four types of items. Rather, certain items were more sensitive to this dimension, irrespective of the child's sex or age. Specifically, the scores on items 2 and 4 were significantly higher than those on items 1 and 3 ($p < .01$ for each of the four comparisons), and the differences between items within these pairs were not significant.

Do items 2 and 4 possess some quality lacking in items 1 and 3 which might account for the former pair's greater sensitivity to the power dimension? Whereas item 1 involves the exchange of an object and item 3 the exchange of meanings with reference to an objective standard of performance, items 2 and 4 seem to connote an exchange of affect directed toward the interacting persons themselves. If this analysis is correct, then it would suggest that children are more sensitive to personalized "love-oriented" than to "object-oriented" relationships (21) in their age-typing. Perhaps the kinds and/or patterning of rewards and punishments associated with love-oriented socialization techniques provide the child with greater incentive to learn age-role discriminations. This interpretation should be viewed with caution, however, for several reasons. Little is known about the differential incentive values of the rewards associated with each of the four types of items. Further, in view of the importance of achievement in middle-class society, particularly in boys, it is surprising that item 3 did not have stronger affective implications. Also, item affect saturation may vary among respondent populations, so that specification of affect criteria without reference to these populations would be misleading. It might be expected, for example, that item 1 would have stronger affective implications in a lower-class sample.

The failure of this study to confirm Parsons' hypothesis of developmental stages in children's age-role concepts should not be viewed as strong evidence against the theory. Parsons' discussion of levels (18) is not sufficiently detailed to be certain that the items used are highly representative of their respective level universes. Also, item differences on other dimensions, such as affect, may have masked the predicted interaction. Finally, since the hypothesis refers to changes in role perception which presumably begin during the first year, the present study may not have sampled age groups which are young enough for the effect to appear.

Discrimination of Age Roles by Attitudinal Direction

The median of the whole sample on the attitude age-role discrimination scale did not differ from chance. There were no significant variations from chance within any of the 12 subgroups, either for the separate mother-girl and father-boy comparisons or for these comparisons combined. It would appear that attitudinal direction is not a dimension in children's age-role discriminations. It should be noted, however, that the power and attitude discrimination scores for the same pair are not completely independent. The fact that the children used the power dimension so intensively and extensively made it less likely, though not impossible, for significant age-role discriminations to occur on the attitude dimension. This situation limits the interpretation, since an independent measure of age-role discrimination by attitude might have revealed trends masked by the power dimension. Nevertheless, if some subjects had relied heavily upon the attitude dimension, or if many had used this dimension a little more frequently, significant intragroup consensus or group differences could have appeared on this as well as on the power dimension. Therefore, it is reasonable to conclude that power is at least relatively more pertinent than attitude as a dimension of age-role discrimination during this age period.

Discrimination of Adult Sex Roles

The whole sample assigned more high power items to the father than to the mother ($p < .001$), and, as seen in Figure 2, there was a curvilinear relation between this tendency and age ($p < .05$). There was no interaction between sex and age on this measure. With respect to the attitude dimension, the median of the whole sample for mother-father discrimination did not differ from chance, and there was no significant variation with age. Although intragroup consensus on this measure was not significant for either sex, there was a difference between sex groups. Girls perceived the mother as somewhat more positive than the father, whereas boys perceived the converse ($p < .05$).

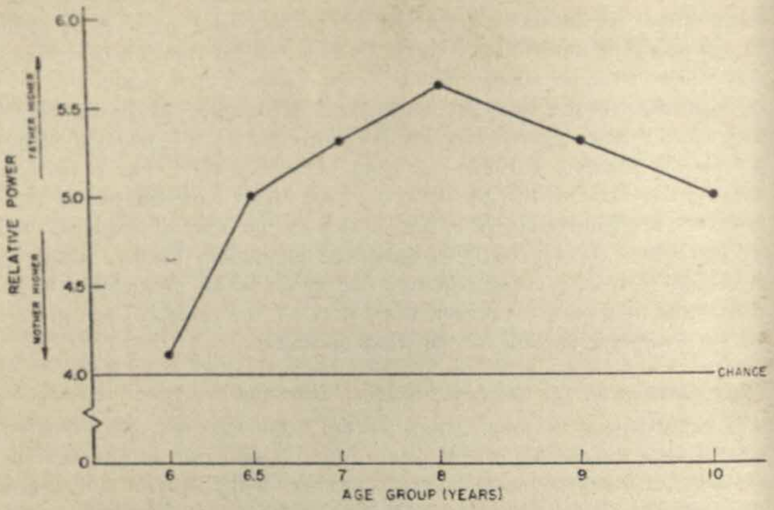

FIGURE 2—Discrimination of parent sex roles by power.

It is most striking that, for the sample as a whole, consensus again occurred on the power dimension, and not on the dimension of attitudinal direction. This finding, together with those on age-role discrimination, indicates that power is the more basic dimension in children's discriminations of parent roles, at least during this age period. If this conclusion is correct, then several interesting implications might follow. Although the importance of power has been recognized in the analysis of intrafamilial relationships (2, 3, 12, 14, 17, 18, 19, 22), many investigators, including the present writer, have assumed that the attitude dimension is at least equally important in parent role differentiation (4, 6, 7, 13, 14, 17, 18, 21). In view of the present results, the latter assumption seems questionable, at least for children of this age. It should be noted, however, that this conclusion does not imply lack of salience of the attitude dimension at other levels, such as that of *person perception,* but only that this dimension may not be pertinent at the level of *role consensus.* Indeed, other results of the present study suggest that several levels need to be considered in a theory of the perceptual factors which guide interpersonal behavior in the child. For example, consider the findings that both sexes perceive the father's sex role as more powerful than the mother's but that girls perceive the mother as more positive, whereas boys perceive the father as more positive. Heider's analysis of benefit and sentiment (11) suggests that a powerful person is seen as benevolent if he is also perceived as positive or as malevolent if he is also perceived as negative; whereas a nonpowerful person is differentiated less sharply in terms of potential benefit. In the present example, it might follow that the father's sex role is perceived as more benevolent by boys and as more malevolent by girls, but that the mother's sex role is not perceived by either sex so clearly in terms of potential benefit or harm.[6]

A second implication is methodological and points out the value of deriving measures from a framework that explicates the various components of family interaction. By distinguishing two possible role dimensions (power and attitude) and two possible position dimensions (sex and age), it became possible to isolate critical role dimensions within specific intrafamilial dyads. Perhaps the often conflicting findings in this area, and consequent difficulties of interpretation, are partially due to inadvertent confounding of these and other interaction components.

The amount of power attributed to the father increased with age between the sixth and eighth year in both the father-boy and mother-father comparisons (Figures 1 and 2, respectively), indicating that this period is particularly significant in the development of father role concepts. Comparison of these two curves also suggests that while there is generalization across the sex- and age-role components of the father role during this period, differentiation of these role components begins to occur between the eighth and tenth year.

The pattern of results for the mother-father comparison differs in several respects from that found in the study of subjects ages 4½ to 6 (7). The younger sample as a whole discriminated parent sex roles on the attitude dimension, whereas the power dimension was used in the present older sample. The obvious interpretation of this difference is that there is a dimension crossover at about the age of 6. The fact that the 6-year-olds of the present study did not discriminate the mother-father pair on the power dimension is consistent with this interpretation. However, this group did not discriminate parent sex roles by attitude either. Further, the girls of the younger sample perceived the mother as more powerful than the father. It is difficult to account for these discrepancies on any basis other than age, since the procedures of the two studies were very similar and the subjects were selected from the same school in successive years. However, in view of the potential importance of a dimension crossover at about the age of 6, it would be desirable to replicate this aspect of the studies in a group ranging in age from 4 to 8 years.

It has been assumed throughout this discussion that the mother-father discrimination measure was derived independently of the measures of age-role discrimination discussed earlier. However, if upon presentation of the mother-father pair, the child imaginatively placed himself into the situation, then this measure could reflect differential parental behaviors toward the child, rather than toward each other. That such confounding may have occurred was suggested by the (nonsignificant) positive association found between amount of power attributed to the father in the mother-father and father-boy comparisons. However, other evidence indicates that this type of confounding did not occur extensively, if at all. First, about a dozen subjects were asked, "Who does he (she) say that to?" after the child responded to a mother-father comparison. In every instance, the child reported that the parent "said" the item to the other parent, not to an imaginatively introduced child. Second, although children perceived the father as significantly more powerful than the mother on the mother-father comparison, they did not attribute greater power to the father in the father-boy comparison than to the mother in the mother-girl comparison. Third, attribution of power to the father in the mother-father comparison was positively correlated with attribution of power to the mother in the mother-girl comparison ($p < .05$), a result opposite to that expected if confounding had occurred. It is also of interest that the mother-girl and father-boy comparisons were positively correlated ($p < .001$). It would appear that attribution of higher power to the father's sex role covaries positively with attribution of higher power to adult roles. Thus, rather than reflecting confounding, the pattern of correlations among the three measures indicates that children's use of the power dimension tends to generalize to both the sex- and age-role components of parent roles.

---

[6] This discussion is meant to provide an illustration, not specific hypotheses. Obviously, the findings on age-role discrimination would have to be incorporated into any set of actual predictions.

Discrimination of Child Sex Roles

Girls perceived the girl role as more positive (and less negative) than the boy role ($p < .001$), whereas boys did not consistently discriminate child sex roles by attitude. There were no significant variations with age for either sex. This finding is in essential agreement with that for children of 4½ to 6 years (7), although differences in the patterns of findings in the two studies lead to somewhat different interpretations. In the study of the younger sample, the fact that girls also discriminated adult sex roles on this dimension led to the conclusion that girls had developed sex-role concepts which generalize across the generations. This conclusion is not warranted in the present study, since girls did not discriminate adult sex roles by attitude. Perhaps this discrepancy between the studies reflects a true developmental change. However, for reasons discussed earlier, it would be important to replicate the findings on adult sex-role perception during the years 4 to 8 in order to evaluate this interpretation.

An interesting pattern of sex differences appears when this measure is broken down into its component items. As seen in Table 3, girls attributed more positive (and less negative) content to the girl role on all four items, particularly on items 2, 3, and 4. Boys, however, used the dimension significantly on fewer items, and not always in the same direction. The fact that both sexes attributed greater achievement competence (item 3) to themselves than to their opposite-sex peers is subject to two interpretations. It may signify a true sex difference in consensus on the definition of child sex roles with respect to achievement competence. Or it may reflect the child's wish to attribute greater competence to himself than to another child, irrespective of the other child's sex. The latter explanation seems plausible in view of the achievement pressures placed upon middle-class children of both sexes. However, the same argument fails to account for the fact that boys did not perceive themselves as significantly more positive than girls on any of the other items. Indeed, on item 4, there was consensus within both sexes that girls are more likely than boys to do the "right thing" and to avoid trouble. What is clear from this pattern of findings is that both sexes use the attitude dimension to discriminate child sex roles, but, whereas girls apply it in the same way to a variety of settings, boys apply it more selectively and in a direction that varies among situations.

TABLE 3

SEX ATTRIBUTED POSITIVE CONTENT ON EACH ITEM FOR GIRL-BOY COMPARISONS

	SEX OF SUBJECT			
	GIRLS		BOYS	
Item Number	*Sex Attributed Positive Content*	*p*	*Sex Attributed Positive Content*	*p*
1	Girl	*ns*	Boy	*ns*
2	Girl	$< .01$	Girl	*ns*
3	Girl	$< .02$	Boy	$< .001$
4	Girl	$< .001$	Girl	$< .01$

The whole sample assigned more low power items to the boy's sex role than to the girl's sex role ($p < .02$), and this tendency did not vary significantly with the child's sex or age. This result is difficult to explain, as it contradicts rather well-established facts on sex differences in children's social behavior (24). It is also inconsistent with the finding on adult sex-role discrimination by power. The implausibility of this result raises a question about the validity of the power measure when applied to child sex-role discriminations. It will be recalled that low power items only were included in this measure because it seemed reasonable to use those items typically assigned to a particular generation when assessing discriminations within that generation. Perhaps this procedure resulted in loss of information that would have provided a more plausible finding, although the same procedure led to quite reasonable findings in the case of adult sex-role discriminations. Another possibility is that the power dimension collapses when the low end is applied to sex-role discriminations within the child generation. If this were the case, the measure could still consistently reflect a dimension, but one that differs from the intended dimension of power. Inspection of the low power items suggests that this measure might have reflected self-centeredness, or simply differential social activity levels. It should be noted that the same question can also be raised about the dimensionality of the power measure applied to adult sex-role discriminations despite the fact that the findings on this measure were more plausible. Further study is required to resolve this problem.

SUMMARY

This study investigated the development of children's conceptions of intrafamilial sex and age roles. The family role concepts of 225 middle-class children of 6 to 10 years of age were assessed by means of a modified paired-comparison procedure. The sample was divided into 12 subgroups based upon the child's sex and age. Intragroup and between groups analyses of role consensus revealed the extent of use of the dimensions of relative power (high vs. low) and attitudinal direction (positive vs. negative) to discriminate intrafamilial roles. The major findings were as follows:

1. In the discrimination of parent roles, there was much greater consensus on the power dimension than on the dimension of attitudinal direction.

2. Children discriminated age roles by assigning high power actions to the adult and low power actions to the child.

3. With increasing age, there was an increasing tendency to discriminate male but not female age roles by power.

4. Some age-role relationships were more differentiated by power than others. It was suggested that children age-type love-oriented more than object-oriented relationships.

5. The father's sex role was seen as more powerful than the mother's sex role.

6. There was a curvilinear relation between the child's age and the extent to which the father's sex role was perceived as more powerful than the mother's sex role.

7. Positive correlations were found among the measures of parent sex- and age-role discrimination by power, suggesting that use of the power dimension to discriminate parent roles generalizes across role situations.

8. Girls discriminated child sex roles by assigning positive actions to the girl and negative actions to the boy. In boys, the discrimination of child sex roles by attitude varied according to the specific interaction situation.

REFERENCES

1. ASCH, S. E. *Social psychology.* Prentice-Hall, 1952.
2. BALDWIN, A. L., KALHORN, J., & BREESE, F. H. Patterns of parent behavior. *Psychol. Monogr.,* 1945, 58, No. 268.

3. Barker, R. G., & Wright, H. F. *Midwest and its children.* Row, Peterson, 1955.

4. Brim, O. G., Jr. The parent-child relation as a social system: I. Parent and child roles. *Child Develpm.,* 1957, 28, 343-364.

5. Bronfenbrenner, U. The study of identification through interpersonal perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford Univer. Press, 1958. Pp. 110-130.

6. Emmerich, W. Parental identification in young children. *Genet. Psychol. Monogr.,* 1959, 60, 257-308.

7. Emmerich, W. Young children's discriminations of parent and child roles. *Child Develpm.,* 1959, 30, 403-419.

8. Gardner, L. P. An analysis of children's attitudes toward fathers. *J. genet. Psychol.,* 1947, 70, 3-28.

9. Gross, N., Mason, W. S., & McEachern, A. W. *Explorations in role analysis.* Wiley, 1958.

10. Hartley, R. E., Frank, L. K., & Goldenson, R. M. *Understanding children's play.* Columbia Univer. Press, 1952.

11. Heider, F. *The psychology of interpersonal relations.* Wiley, 1958.

12. Hoffman, M. L. Power assertion by the parent and its impact on the child. *Child Develpm.,* 1960, 31, 129-143.

13. Kagan, J. The child's perception of the parent. *J. abnorm. soc. Psychol.,* 1956, 53, 257-258.

14. Kagan, J., & Lemkin, J. The child's differential perception of parental attributes. *J. abnorm. soc. Psychol.,* 1960, 61, 440-447.

15. Mead, G. H. *Mind, self and society.* Univer. of Chicago Press, 1934.

16. Meltzer, H. Sex differences in children's attitudes to parents. *J. genet. Psychol.,* 1943, 62, 311-326.

17. Mussen, P., & Distler, L. Masculinity, identification, and father-son relationships. *J. abnorm. soc. Psychol.,* 1959, 59, 350-356.

18. Parsons, T. Family structure and the socialization of the child. In T. Parsons, R. F. Bales, & Associates, *Family, socialization, and interaction process.* Free Press, 1955. Pp. 35-131.

19. Radke, M. J. The relation of parental authority to children's behavior and attitudes. *Univer. Minn. Inst. Child Welf. Monogr.,* 1946, No. 22.

20. Sarbin, T. R. Role theory. In G. Lindzey (Ed.), *Handbook of social psychology. Vol. 1.* Addison-Wesley, 1954. Pp. 223-258.

21. Sears, R. R., Maccoby, E. E., & Levin, H. *Patterns of child rearing.* Row, Peterson, 1957.

22. SYMONDS, P. M. *The psychology of parent-child relationships.* Appleton-Century, 1939.

23. TASCH, R. J. The role of the father in the family, *J. exp. Educ.*, 1952, 20, 319-361.

24. TERMAN, L. M., & TYLER, L. E. Psychological sex differences. In L. Carmichael (Ed.), *Manual of child psychology.* Wiley, 1954. Pp. 1064-1114.

25. WARNER, W. L., MEEKER, M., & EELS, K. *Social class in America.* Science Research Associates, 1949.

26. WILSON, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, 53, 96-101.

CHILD'S SYMBOLIC CONCEPTUALIZATION OF PARENTS[1]

JEROME KAGAN,* BARBARA HOSKEN,
and SARA WATSON

Fels Research Institute

The language labels that the child applies to objects and events in his environment are an important determinant of much of his behavior. When the child learns to apply the word "dangerous" to objects with no feet that slither on the ground, his behavior toward this class of stimuli will be a function of the response that he has learned to the label "dangerous"—usually flight. This economical use of language to mediate a variety of behaviors is usually adaptive and allows the person to behave appropriately in a variety of new situations. However, the application of certain labels can lead to less desirable reactions. The semantic link between "disobedience" and "bad" often results in uncomfortable feelings when violation of parental standards has been committed or contemplated. Similarly, the semantic link "woman-flighty" predisposes an individual to react initially to all women as if this equation were universally true.

Parent figures are an important set of social objects with whom the child constantly interacts. The constellation of labels that the child ascribes to each parent will influence not only his reactions to them, but also his behavior with parent substitute figures (teachers, relatives, employers). Since the parents supply the child with his first definition of sex roles, the labels that are applied to father and mother are apt to generalize initially to all males and females.

During recent years there have been several investigations of the young child's conceptualization of parent roles (3, 4, 5, 7, 8). The results of these studies have been in remarkable agreement in indicating that the father, in relation to the mother, is perceived as more punitive, more feared, more

---

* Fels Research Institute, Yellow Springs, Ohio.

[1] This research was supported, in part, by research grant M-1260 from the National Institute of Mental Health, United States Public Health Service.

Child Develpm., 1961, 32, 625-636.

dominating, and less nurturant. This pattern of childhood associations linking father with aggressivity and mother with nurturance is consistent with the semantic differentiation of masculinity and femininity made by young adults (10).

It is generally assumed that the concepts *male* and *female* are linked with descriptive dimensions that are highly symbolic of these concepts. Psychoanalytic writings and anthropological reports suggest that dangerous animals, dark colors, and pointed objects are symbolic of masculinity; domestic animals, light hues, and rounded objects are presumably symbolic of femininity. In a recent empirical study with different cultural groups, it was found that angular lines, dark hues, and large objects were more frequently associated with men; their polar opposites were typically ascribed to women (9). It seemed profitable, therefore, to investigate the degree to which the young child had acquired these highly symbolic representations of sex roles.

The primary purpose of this research was to investigate the child's differential conceptualization of the concepts *mother, father,* and *self* on a variety of symbolic dimensions that are presumably linked both with aggressivity and nurturance, as well as masculinity and femininity. An ancillary goal was to replicate the results of earlier studies (7, 8) that indicated a perception of the father as more feared and more potent than the mother.

METHOD

Subjects and Procedure

*S*s were 98 white children (59 boys and 39 girls) from intact families. Their ages ranged from 6 to 8 years with a modal and median age of 6½ years. All *S*s were attending the first or second grade in one of three elementary schools at the time of data collection. Each *S* was individually interviewed by one of two female experimenters (BH or SW) on three separate occasions. The sessions were separated by a period of 7 to 14 days with a modal intersession interval of 10 days. Several days prior to the first test session the *E* visited the child's classroom in order to become acquainted with each *S* and reduce the effect of stranger anxiety on the child's test performance.

During the first test session, half of the *S*s were questioned about their conceptualization of their father; the remaining half were tested for their conception of mother. During the second session each child was questioned about the second parent. On the third session three series of questions were administered: (a) the child's conception of himself, (b) his conception of the concept "bad," and (c) a picture completion test.

The Tests

Symbolic conceptualization of father, mother, self. The technique used to assess *S*'s conceptualization of *father, mother,* and *self* consisted of 66 pairs of picture stimuli. Each pair illustrated the polar opposite aspects of a descriptive dimension. The 11 dimensions sampled were: strong-weak, big-little, nurturant-nonnurturant, competent-incompetent, punitive-nonpunitive, dangerous-harmless, dirty-clean, dark hue - light hue, cold-warm, mean-nice, and angular-rounded. Each of these 11 dimensions was presented six times with six different pairs of relevant pictures. The pictures contained no illustrations of men or women nor any clue that the child might interpret as directly suggestive of a male or a female.

After *E* had gained the rapport of the child, she said, "I'm going to show you some pictures, and I want you to tell me which one reminds you of your (father, mother, self). Which one is like your (father, mother, self)?"

E then showed the *S* some practice stimuli that were obvious in their appropriateness for men, women, or children (e.g., trousers vs. dress, apron vs. men's shoes, child's toy vs. piece of paper). *E* then administered the 66 test stimuli. *E* read a verbal description of the two pictures on each card that contained the salient symbolic dimension. *E* would say, for example, "Here is a strong rabbit, and here is a weak rabbit. Which one reminds you of your (father, mother, self)?" or, "Here is a mouse that knows how to read, and here is a mouse that does not know how to read. Which one reminds you of your (father, mother, self)?"

The presentation of the 11 dimensions was randomized as was the position of the pictures in each pair. *E*'s verbal descriptions of the 66 pairs of pictures follow:

1. Strong rabbit and weak rabbit.
2. Little boat and big boat.
3. Someone giving money to child and someone not giving anything.
4. Mouse that knows how to read and mouse that does not know how to read.
5. Someone who is mad at child and someone who is not mad at child.
6. A lamb and a lion.
7. A clean dog and a dirty dog.
8. Black horse and a white horse.
9. Cold bowl of soup and a warm bowl of soup.
10. Mean whale and a nice whale.
11. Round shape and a pointed shape.
12. Weak piece of wood and strong piece of wood.
13. Big ball and little ball.
14. Someone who does not want to pet the dog and someone who likes to pet the dog.
15. Bird who is lost and doesn't know where to go and bird who does know where to go.
16. Someone not slapping child's hand and someone who is slapping child's hand.
17. A tiger and a horse.
18. Dirty foot and a clean foot.
19. White bear and black bear.
20. Warm fireplace and cold fireplace.
21. Nice cat and a mean cat.
22. This shape and that shape.

23. Strong fence that stood up and weak fence that fell down.
24. Little table and big table.
25. Someone covering the cat so it will be warm and someone not covering the cat.
26. Dog that can do tricks and a dog that can not do tricks.
27. Someone telling child to go to (his) (her) room and someone not doing this.
28. A hammer and a gun.
29. Clean room and a dirty room.
30. Brown table and a red table.
31. Cold day and a warm day.
32. Mean squirrel and a nice squirrel.
33. This line and that line.
34. Strong plate that did not break and weak plate that did break.
35. Big block and a little block.
36. Someone not kissing the baby and someone kissing the baby.
37. Squirrel that doesn't know where to look for food and a squirrel that does know where to look.
38. Someone not yelling at child and someone yelling at child.
39. Crocodile and a bird.
40. Dirty pig and a clean pig.
41. White telephone and black telephone.
42. Warm sweater and a cold sweater.
43. Nice cow and a mean cow.
44. This design and that design.
45. Strong rope and a weak rope.
46. Little telephone and big telephone.
47. Someone bringing the lamb into the house and someone telling the lamb to go away.
48. Pitcher that will pour the right way and pitcher that won't.
49. Someone scolding the child and someone not scolding the child.
50. A rabbit and a snake.
51. Clean ash tray and a dirty ash tray.
52. Purple hat and a yellow hat.
53. Cold bed with thin blanket and a warm bed with a thick blanket.
54. Mean elephant and a nice elephant.
55. This line and that line.
56. Weak cloth that tore and strong cloth that did not tear.
57. Big cup and little cup.
58. Someone who will not help calf out of the wire and someone who will help the calf get loose.
59. Clock that doesn't remember to ring and a clock that does remember.
60. Hand that will not spank the child and a hand that will spank the child.
61. A knife and a ruler.
62. Dirty bed and a clean bed.
63. Light blue pencil and dark blue pencil.
64. Warm bath and cold bath.
65. Nice mouse and a mean mouse.
66. This shape and that shape.

This set of 66 stimuli was administered on each of the three sessions to assess the child's conception of father, mother, and self. After *E*'s verbal description of each stimulus, *S* pointed to the picture he felt reminded him of his father, mother, or self.

The concept of "bad." On the third test session *S* first went through the 66 stimuli indicating which picture best applied to himself. *E* then asked the child if he knew the meaning of the word "bad." When *E* was satisfied that the child was familiar with the word she said, "Now we are going to look at these pictures again and I want you to tell me which one is bad. Which picture is the bad one?" *E* then administered the first 44 stimuli.

Picture completion: Perception of parents. The final test in the third session consisted of 12 pictures (separate pictures for boys and girls) in which a boy or girl was illustrated in a different situation but no adult was illustrated. *E* said, "Here are some pictures which show a (boy, girl) but the mommy or daddy is missing. You tell me who is missing, the mommy or the daddy."

The pictures were designed to suggest parental nurturance, punitiveness, or strength. *E*'s verbal description for each of the 12 pictures follows:

PICTURE 1. Someone is giving the (boy) (girl) an ice cream cone. Who is it, the mommy or the daddy?

PICTURE 2. Someone is going to get the (boy) (girl) the toy (he) (she) can't reach. Who is getting it, the mommy or the daddy?

PICTURE 3. Someone just spanked the (boy) (girl). Who spanked (him) (her), the mommy or the daddy?

PICTURE 4. Someone is picking the (boy) (girl) up in the air. Who is doing it, the mommy or the daddy?

PICTURE 5. Someone is yelling at the (boy) (girl). Who is yelling, the mommy or the daddy?

PICTURE 6. Someone is giving the (boy) (girl) a present. Who is giving the present, the mommy or the daddy?

PICTURE 7. The dog is chasing the (boy) (girl) and (he) (she) is running to someone for help. Who is the (boy) (girl) running to, the mommy or the daddy?

PICTURE 8. The (boy) (girl) is reading a book with someone. Who is (he) (she) reading with, the mommy or the daddy?

PICTURE 9. Someone is going to spank the (boy) (girl). Who is it, the mommy or the daddy?

PICTURE 10. Someone is taking a walk with the (boy) (girl). Who is it, the mommy or the daddy?

PICTURE 11. Someone is on the seesaw with the (boy) (girl). Who is it, the mommy or the daddy?

PICTURE 12. The (boy) (girl) is mad at someone. Who is (he) (she) mad at, the mommy or the daddy?

RESULTS

Symbolic Conceptualization of Parents and Self

For convenience of presentation only one pole of each of the 11 dimensions will be considered. Since *S* had to choose between the two poles, the scores for either pole accurately reflect the nature of the results. The 11 attributes to be discussed are strong, big, nurturant, competent, punitive, dangerous, dirty, dark, cold, mean, and angular.

The data were initially analyzed for differences with respect to experimenter, school, and order of questioning (father or mother in first session). None of these differences even approached statistical significance, and the data for experimenters, schools, and order of questioning were pooled in the presentation to follow. Table 1 presents the average score on the 11 dimensions for father, mother, and self for the sexes separately and for the total group.

TABLE 1

AVERAGE SCORE ON CONCEPTUALIZATION OF FATHER, MOTHER, AND SELF

	FATHER			MOTHER			SELF		
Dimension	*Boys*	*Girls*	*Total*	*Boys*	*Girls*	*Total*	*Boys*	*Girls*	*Total*
Strong	5.6	5.9	5.7	4.7	5.1	4.9	5.6	4.8	5.4
Large	5.7	5.8	5.7	4.1	4.1	4.1	4.9	3.5	4.4
Nurturant	5.2	5.2	5.2	5.0	5.6	5.3	5.2	5.5	5.3
Competent	5.4	5.2	5.3	5.2	5.3	5.2	5.4	5.2	5.3
Punitive	0.9	0.9	0.9	0.8	0.4	0.6	0.7	0.6	0.7
Dangerous	1.8	1.7	1.8	0.6	0.4	0.5	1.8	0.4	1.3
Dirty	0.6	0.7	0.7	0.2	0.1	0.2	0.7	0.5	0.6
Dark	3.0	3.0	3.0	2.0	1.8	1.9	2.7	1.7	2.3
Cold	0.6	0.5	0.6	0.8	0.3	0.6	1.1	0.8	1.0
Mean	0.8	0.9	0.8	0.5	0.1	0.4	0.8	0.6	0.6
Angular	2.7	2.0	2.4	1.9	1.3	1.6	2.2	1.2	1.8

Differential conception of parents. Boys and girls agreed that father, in relation to mother, was stronger, larger, more dangerous, more dirty, darker, and more angular. Each of these differences between maternal and paternal attributes was significant at the .01 level or better (all p values are for two tails). Girls described the father as more punitive and meaner than the mother ($p < .05$), but differences on these two dimensions were not significant for boys. There were no significant differences between the parents for the attributes nurturant, cold, or competent.

Conception of self. Boys, in comparison to girls, labeled themselves as stronger, larger, more dangerous, darker, and more angular. Each of these differences was significant at the .01 level or better. These five dimensions were also more frequently ascribed to the father than the mother by both sexes. Thus, boys and girls conceptualized themselves as more similar to the same than the opposite sex parent on the dimensions strong, large, dangerous, dark, and angular.

In order to obtain a more direct measure of the similarity between the child's conceptualization of himself and the same sex parent, difference scores were computed for each dimension between the score for each parent and the score for the self. For example, the number of times the child labeled himself as "strong" was subtracted from the number of times he labeled his father or mother as "strong." These differences between the scores for the self and each parent were computed for all 11 dimensions. The mean difference between father and self was larger for girls than for boys ($p < .05$); the mean difference between mother and self was larger for boys than for girls ($p < .05$). Thus, the 6-year-old child labels himself as more similar to the same than to the opposite sex parent on attributes that are highly symbolic of the sex roles.

TABLE 2

AVERAGE "BAD" SCORE FOR EACH DIMENSION

Dimension	*Boys*	*Girls*	*Total*
Strong	0.8	0.6	0.7
Large	1.5	1.3	1.4
Nurturant	0.3	0.1	0.2
Competent	0.5	0.5	0.5
Punitive	3.8	3.8	3.8
Dangerous	3.9	4.0	3.9
Dirty	3.8	3.8	3.8
Dark	2.7	2.7	2.7
Cold	3.3	3.6	3.4
Mean	4.0	3.9	3.9
Angular	2.9	3.0	2.9

Conception of the Word "Bad"

Table 2 presents the average number of times a dimension was called "bad" by the child (range 0 to 4).

Boys and girls were again in close agreement that the dimensions, mean, dangerous, dirty, punitive, cold, weak, nonnurturant, and incompetent, were "bad" (all had average scores over 3.0). The high degree of conceptual agreement between boys and girls is reflected in the rank-order correlation of +.99 for the differential assignment of the word "bad" to the 11 dimensions.

Picture Completion Test for Perception of Parents

Table 3 presents the percentages of boys and girls responding mother or father to each of the 12 pictures.

These data agree with those obtained with the more symbolic stimuli. On pictures 3 and 9, which illustrate physical punishment of the child, boys and girls agreed that father was more likely to be punitive than

Table 3

PERCENTAGE OF GROUP RESPONDING MOTHER OR FATHER TO THE TWELVE PICTURES

	Boys		Girls		Total	
Picture	*Father*	*Mother*	*Father*	*Mother*	*Father*	*Mother*
1. Child receiving ice cream	54.4	43.8	26.3	71.0	43.2	54.7
2. Person reaching for toy	63.2	36.8	55.3	42.1	60.0	38.9
3. Child spanked	63.2	36.8	68.4	31.6	65.3	34.7
4. Child picked up	84.2	15.8	94.7	5.3	88.4	11.6
5. Child yelled at	43.8	54.4	50.0	50.0	46.3	52.6
6. Child receiving present	22.8	75.4	23.7	68.4	23.2	72.6
7. Child fleeing dog	66.7	31.6	52.6	47.4	61.0	37.9
8. Child read to	42.1	57.9	23.7	76.3	34.7	65.3
9. Child to be spanked	71.9	28.1	71.0	28.9	71.6	28.4
10. Child walking	33.3	66.7	28.9	71.0	31.6	68.4
11. Child on seesaw	70.2	22.8	28.9	57.9	53.7	36.8
12. Child mad	50.9	45.6	71.0	26.3	58.9	37.9

mother ($p < .01$). On pictures 2, 4, and 7, which illustrate scenes requiring strength, size, and the ability to protect the child from danger, boys and girls both answered father more than mother ($p < .05$). To pictures 6 and 10, which illustrate nurturance and companionship, *S*s responded mother more than father ($p < .01$). Responses to these pictures indicate that strength and punitiveness were viewed as more characteristic of the father while nurturance was ascribed to the mother.

Answers to some of the pictures yielded sex differences. To pictures 1, 8, and 11, which suggest nurturance, girls were more likely to reply mother than the boys ($p < .02$, .10, and .01 respectively). To picture 12 (child mad at someone) girls were more likely to answer father than were boys ($p < .05$). Thus, girls were more likely than boys to view the mother in a nurturant light and the father in a hostile one. This pattern agrees with the girls' responses to the 11 symbolic dimensions. Girls labeled the father as more punitive and meaner than the mother; boys did not make this distinction between the parents.

Discussion

Most of these results are in essential agreement with previous findings that were based on different samples of children and different procedures (3, 4, 5, 7, 8). The child views the father as more punitive, more fear arousing, more potent, but less nurturant than the mother.

Responses to the symbolic stimuli did not indicate that nurturance was more characteristic of the mother. This result seems due, in part, to the procedure used with this instrument. In previous research, the child was presented with a description of a nurturant act or object and *forced to choose between mother or father* as the more appropriate parent for the stimulus. Under these conditions the child typically chose the mother. In the procedure used in this study, the child was not forced to choose between mother or father and was allowed to label both parents as nurturant. This is exactly what these *S*s did. The stimuli used to illustrate the nurturant dimension forced the child to pick between a nurturant event and one characterized by an absence of nurturance. All of the children agreed that lack of nurturance was a "bad" characteristic, and they were reluctant to assign a "bad" trait to either parent. The picture completion test, which used the more popular procedure of forcing the child to pick between mother or father, yielded results that were in agreement with previous investigations. The mother was labeled as more nurturant than the father on pictures 6, 8, and 10.

The present data verify some of the classical hypotheses regarding symbolic manifestations of sex roles. Angularity of line, dark hues, dangerous animals, and large objects appear to be symbolic of the father and of masculinity in the conceptual organization of the school age child. The work of Jenkins and Russell (6) and Osgood (9, 10) indicate the similarity in conceptualization of sex roles between children and adults. When college students rated the concepts of father and mother on the semantic differential, the father received higher scores than the mother on the dimensions, cruel, successful, strong, and bad (6). A recent study (9) of Anglo, Japanese, Navajo, and Mexican-Spanish adults used a procedure similar to the one used in this research and yielded similar results. *S*s were presented with 13 pairs of visual stimuli that signified dimensions such as verticality, thickness, angularity, size, and darkness. *S* matched one of these pairs of stimuli with concepts such as man, woman, bad, heavy, good, fast, happy, and strong. The results were in essential agreement with those found with the present group of children. For example, the dark stimulus was labeled "bad" and was more often associated with "man" than with "woman" by all four language-culture groups. Similarly, the rounded line design, as opposed to the angular design, was more frequently assigned to the concept "woman" than to the concept "man." Largeness, on the other hand, was assigned more often to "man" than to "woman." Thus, there appears to be a universality surrounding the dimensions angular, large, and dark that applies to children and adults as well as to people from different language cultures. It seems fair to conclude that some of the labels that are implicitly linked with the concepts male and female in the adult psyche have been acquired by the time the child is 5 or 6 years old.

The language learning that accounts for this communality in the conceptualization of sex roles probably occurs very early, perhaps as soon as language becomes available to the child. The degree of homogeneity between child and adult and across language cultures suggests that factors outside the child's immediate family are partly responsible for this language learning. All fathers are not more punitive or more dangerous than mothers, yet most children respond as if this were so. It is necessary to turn to the consequences of certain universal experiences as well as to the effect of the mass media in order to explain this phenomenon.

In western culture the major sources of communication portray the male role as markedly more aggressive and violent than the female role, and most adults concur in this evaluation (1, 2, 6). It is reasonable to expect, therefore, that animals and objects that are usually linked with aggression and danger will be associated with masculinity. At a point further out on the generalization continuum, it is plausible to assume that any stimulus dimension that is associated with pain, danger, or anxiety will be preferentially assigned to the male role. Some of our children verbalized that the "angular" designs can "hurt you," and the association between pointed objects (needles, nails, knives) and pain may account for the association matrix linking angularity with masculinity. The association between "darkness" and "father" could be a function of the more frequent use of dark clothes by men as well as the association cluster that includes darkness, the night, and danger.

It is possible that this generalized perception of men as more dangerous and stronger than women influences each child's perception of his own father. That is, once the child has acquired the association between danger and men and nurturance and women, he may begin to anticipate aggression from his father and affection from his mother. The content of these sex role concepts can also affect an individual's initial posture with strange males and females. If child and adult agree that boys and men are stronger, larger, and more dangerous than girls and women, an initial attitude of timidity and passivity will be more likely to prevail with strange men than with strange women.

It is interesting to note that on both the symbolic stimuli and the incomplete pictures, girls labeled the father as more hostile and punitive than did boys. This same finding emerged with an independent sample of 67 children who averaged one year younger than the present sample (8). The replication of this result reinforces the conclusion that the young girl views the father with more fear than the young boy does. It is suggested that, because the 6-year-old boy has begun to identify with his father, he is forced to repress some of the father's fear arousing qualities.

The Evaluative Dimension

The attributes, strong, large, nurturant, competent, clean, warm, nice, and nonpunitive, were most frequently assigned to both parents. The opposite poles of weak, small, nonnurturant, incompetent, dirty, cold, mean, and punitive were infrequently assigned to either parent. These latter attributes were labeled "bad" by the child. It appears that the child's reluctance to assign "bad" characteristics to his parents was a major determinant of his answers to the 66 symbolic stimuli. Rank-order correlations between the average "bad" score for each dimension and its frequency of assignment to the parents were —.75 and —.94 for father and mother, respectively. Thus, the assignment of traits to the parents was regulated, in large measure, by the child's evaluation of the negative qualities of that dimension.

This pervasive influence of a good-bad dichotomy on the child's responses agrees with Osgood's conclusion that the evaluative dimension is one of the primary structuring principles that people use in classifying people, objects, and events (10).

SUMMARY

A group of 59 boys and 39 girls, age 6 to 8, were interviewed on three separate occasions for their conceptualization of their parents and themselves. A special instrument was devised to assess the child's conceptualization of father, mother, and self on 11 polar dimensions: strong-weak, big-small, nurturant-nonnurturant, competent-incompetent, punitive-nonpunitive, dangerous-harmless, dirty-clean, dark-light, cold-warm, mean-nice, and angular-rounded. In addition, the child was asked to supply the missing figure (father or mother) to each of 12 scenes illustrating children in different situations.

Boys and girls conceptualized the father, in comparison to the mother, as stronger, larger, darker, more dirty, more angular, and more dangerous. The children conceptualized themselves as being more similar to the same sex than to the opposite sex parent. Answers to the incomplete pictures revealed that the mother was labeled as more nurturant but less punitive and less potent than the father.

It was suggested that the child's contact with certain common, environmental contingencies and communications from the public media were important factors in the symbolic association between masculinity and the dimensions of angularity, darkness, size, strength, and danger. The semantic domains of the concepts male and female held by the 6-year-old overlap in large measure with the conceptualizations of the adult.

REFERENCES

1. BENNETT, E. M., & COHEN, L. R. Men and women: personality patterns and contrasts. *Genet. Psychol. Monogr.*, 1959, 60, 101-153.
2. CHILD, I. L., POTTER, E. H., & LEVINE, E. M. Children's textbooks and personality development: an exploration in the social psychology of education. *Psychol. Monogr.*, 1946, 60, No. 279.
3. EMMERICH, W. Young children's discriminations of parent and child roles. *Child Develpm.*, 1959, 30, 403-419.

4. Finch, H. M. Young children's concept of parent roles. *J. Home Econ.*, 1955, 47, 99-103.

5. Gardner, P. L. An analysis of children's attitudes toward fathers. *J. genet. Psychol.*, 1947, 70, 3-28.

6. Jenkins, J. J., & Russell, W. A. An atlas of semantic profiles for 360 words. *Amer. J. Psychol.*, 1958, 71, 688-699.

7. Kagan, J. The child's perception of the parent. *J. abnorm. soc. Psychol.*, 1956, 53, 257-258.

8. Kagan, J., & Lemkin, J. The child's differential perception of parental attributes. *J. abnorm. soc. Psychol.*, 1960, 61, 440-447.

9. Osgood, C. E. The cross-cultural generality of visual-verbal synesthetic tendencies. *Behav. Sci.*, 1960, 5, 146-169.

10. Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning*. Univer. of Illinois Press, 1957.

ASSESSING SOME ASPECTS OF THE PARENT-CHILD RELATIONSHIP

F. N. Cox* *and* Patricia M. Leaper
University of Melbourne

In this study we wished to make a systematic examination of some selected aspects of parent behavior on the adjustment of a diverse, yet typical group of Australian primary school children. In selecting the aspects of parents' behavior we would study, we decided to rely largely on propositions formulated by psychoanalytically-oriented clinicians and anthropologists. We felt it would be an advantage, however, to assume that certain components of syndrome complexes of parent behavior can be isolated and studied separately. It was thought that such an assumption would simplify the process of evolving operational definitions of the aspects of the parent-child relationship we chose to study.

Theoretical Considerations

The first problem, then, was that of choosing which of the many aspects of parent behavior we would examine. As has been implied, selection of variables was strongly influenced by the heuristic value of psychoanalytically-oriented propositions. Nearly all writers with this point of view have emphasized the supreme importance of "love" for child adjustment. "Restriction" is also frequently stated to have significant effects, while "family cohesion" or "family harmony" is another factor which some of these writers claim can be important for understanding children's adjustment.

Love

In our thinking about this variable we were particularly influenced by the writings of Anna Freud (7, 8), Erikson (6), Jersild (11), and Balint (3). From certain of their propositions, we tried to evolve a definition of "love" which would serve as a basis for constructing measuring instruments. More especially, we attempted to derive a definition which would delineate a universe of behaviors which constitute "love," and to do this in such a

---

* Department of Psychology, University of Melbourne, Melbourne, Victoria, Australia.

Child Develpm., 1961, 32, 637-649.

way that we could then construct items to sample this universe.

The generalized definition of "love" adopted in this study is that the parents' behavior should be expressive of a set of attitudes which we term child oriented, developmentally relevant, and child respecting. The *child oriented* attitude is expressed in using childrearing methods that take account of what the child actually needs, as well as what the parents think he needs, and what they might want him to need. An important implication is that parents do not repeatedly use their child as a means of gaining narcissistic satisfactions.

The *developmentally relevant* attitude is expressed by parents setting expectations and demands that are appropriate to the emotional, physical, and social capacities and skills of the child at a given stage of his development. In order to ensure that the child will not be consistently forced to cope with insoluble problems, parents will set definite and relevant goals for him. This implies that parents will have actively familiarized themselves with what can be reasonably expected of children at various ages.

The *child respecting* attitude is expressed by parents showing appreciation of the child's individuality. They recognize and endorse the child's emotional, intellectual, and physical strengths and limitations, including his susceptibility to urgent, internal stimulation. They also grant their child some autonomy and privacy.

It is apparent that this type of definition excludes a number of notions concerning "love" that psychoanalytic and other writers have stressed. For instance, it is not a definition which emphasizes demonstrativeness or warmth; it does stress, however, the importance of well-informed guidance. Moreover, this definition differs in detail from the propositions of any single writer in this field, but we thought that it would delimit a universe of behaviors from which a set of items for assessing variations in parent behaviors could be derived. The nature of the items would vary from one age level to another, and items at any age level would have to be related to the known facts of development at that age.

Restriction

The importance of control, limit-setting, and overprotection have been discussed by such writers as Baldwin (1), Erikson (6), Anna Freud (7), Jersild (11), Levy (12), Lewin (13), and Redl and Wineman (15). We attempted to derive a definition of "restriction" which covered as many of the propositions of the above workers as could be synthesized into a coherent pattern. That is, we tried to delineate a unitary but broad universe of parental behaviors, ranging from neglect through control (limit-setting) to overprotection. While this approach may not lend itself to covering all of the extremes of overindulgence and overrestriction described by Lewin and Levy, it does have the advantage of making an a priori distinction between "restriction" and the above propositions referring to the guidance aspects of "love." We defined our universe of "restriction" as consisting in those parental behaviors which confine children's freedom of action. It was argued that degrees of restriction could be measured by assessing the extent to which parents demand that their child participate in household routines, confine his activities to specified locations, and associate only with designated persons. While failure to supervise any of the child's activities probably implies neglect, overrestriction is implied when parents fail to ensure that their child has some time to himself, chances to explore his neighborhood, and opportunities for mixing with people living around him.

Family Cohesion

Various terms, such as atmosphere, integration, cohesion, and harmony, have been used to describe the family as a group. The variety of these terms indicates that the group functions of a family can be looked at from different points of view. It is apparent that the different points of view arise because of different predictive purposes.

For instance, it seemed to us that the propositions advanced by Lewin (14) and Baldwin (2) differ from those advanced by writers such as Erikson (6) and Friedlander (9), in that Lewin and Baldwin have not been concerned with differentiating symbiotic from other family relationships. For the limited purposes of the present study, we decided to define family cohesion in terms of the degree to which the various functions necessary for the maintenance of the family unit are apportioned according to mutually agreed preferences and skills. We argued that these family functions consist of providing and managing finance, food, and shelter; initiating and consolidating the social relationships of the family unit to the extended family, the neighborhood, and the wider community; and providing facilities and opportunities for constructive use of leisure time. We decided to use measures of mother, father, and child joint participations in the functions of initiating and consolidating social relationships and constructive use of leisure time as indices of family cohesion, for we intended to use children's reports on family activities. It seemed improbable that children would be able to give information about the nature of the division of sustenance functions within the family. Further, even in those cases in which children are capable of observing the division of sustenance functions they may misinterpret them, particularly if the division is so well established that discussion about it rarely focuses upon central issues.

Family cohesiveness will be expressed when the parents join with each other to visit and entertain, enjoy casual leisure activities in the home, and go out on family excursions and include their children whenever their activities are appropriate.

The variables of love, restriction, and cohesion comprise those aspects of the parent-child relationship we chose to study in detail. Many people will argue that other significant factors could be profitably substituted for our variables. In particular, many will claim that "consistency" of parental actions is of paramount importance for predicting child adjustment. However, it appears to us that consistency can be regarded in two ways: either as a modifier of other variables or as a factor having independent effects on child behavior. We did not feel at all confident that the methodology to be described below would be appropriate to a serious study of consistency and its effects, for the study relies on children's accounts of activities in the home.

METHODOLOGY

Basic Assumptions

The methodology adopted in this study rested upon two interrelated assumptions: namely (a) that 10- to 11-year-old children can give as accurate an account of certain specified aspects of parent behavior as parents can and (b) that children of this age are sufficiently mature to report what usually happens without merely reflecting the events of a particular day.

Population

To initiate our studies of whether the variables of Love, Restriction, and Family Cohesion could be measured by administering structured question schedules to children, we decided to begin with a restricted population. The population specifications were that (a) the children should be boys in the fifth grade at primary (elementary) school and, therefore, the great majority would be 10 or 11 years of age; (b) the sample should be drawn from a wide cross-section of the local urban community; and (c) the children should, if possible, come from intact family units and be living at home.[1]

These specifications were set down to ensure that the children were capable of writing answers to clearly worded questions;[2] that changes in development associated with puberty would not complicate the initial attempts to measure the family variables; and that the extrafamilial influences to which the children were exposed were as uniform as possible—for example, the broad course of their school activities would be approximately equivalent. Further, we wanted to concentrate on the construction of measures which would, in the first instance, distinguish parental behaviors in "normal," sociologically intact family units.[3]

Secondary Assumptions

The basic assumption that the specified population of children will give accurate information about parental behavior will only obtain if the information is sought under the following conditions: (a) if questions relevant to a given variable sample a wide variety of the situations in which that variable is likely to operate; (b) if children have had frequent and repeated opportunities to observe the parental behaviors about which they are questioned; (c) if questions are directed toward eliciting statements of fact rather than of opinion; (d) if questions are framed in such a way that they do not arouse undue emotional reactions. Emotionally provocative questions would have an unknown effect upon the veracity of answers to the question itself and to those that follow it, quite apart from stimulating unnecessary and possibly uncontrollable anxieties in the children.

---

[1] These studies were carried out in the city of Melbourne, the population of which is approximately 1,750,000.

[2] Mentally defective children and migrant children who cannot understand and write English were deliberately excluded.

[3] It is apparent that this specification is designed to exclude institutionalized children and children who do not live with both parents.

It was soon apparent that no existing questionnaire appropriate to the specified population was designed so that these conditions obtained. However, perusal of existing questionnaires suggested a particularly useful technique for the construction of question schedules—the daily life schedule. It seemed that this technique would simplify presentation of provocative questions, and so help to gain children's acceptance of, and cooperation in, the task. Further, it was thought that the use of the daily life schedule technique would facilitate the adequate sampling of situations in which the Love, Restriction, and Family Cohesion variables would be likely to operate. Presentation of separate items in a time sequence also has the effect of dispersing items relevant to a particular variable. Finally, the technique enables interpolation of "buffer items" with face validity, which in itself would seem to help to divert attention from, and reduce reaction to, some necessarily emotionally potent items.

Form and Content of Scales

All of the Love items consisted of questions about situations in which sympathetic and discreet parental participation is necessary for the optimal psychological well-being of a 10- to 11-year-old child. Specifically, children were asked if parents exercise firm but sympathetic control over the ways in which they, the children, begin and end their day; if, and how, parents provide opportunities for communication about the children's reactions to pleasurable and unpleasurable experiences; and about the quality of parental reactions to especially good and bad child behaviors.

The form of the items intended to measure Love varied. Most consisted of a related series of open-ended and closed questions, the former requiring only short answers. For example, the form of an item about the quality of parental reactions to especially good child behaviors was: (a) "If you do something especially good, does someone in your family praise you for doing it?" (b) "Put a ring or circle around the person or people who praise you." M F B S b s O.[4] (c) "How do these people praise you?" Scoring of such items was necessarily interpretative and depended upon the patterns of answers to the series of questions comprising one item. The remainder of Love items consisted of single multiple-choice, closed questions.

---

[4] The symbols M, F, B, S, b, s, O designated mother, father, elder brother, elder sister, younger brother, younger sister, and other adults living with the family, respectively. These symbols were used in many items and the children had no difficulty with them.

Specific items in the Restriction scale elicited information about the extent to which a child participates in household routines, has to confiine his activities to specific locations, and has to restrict his associations with other people. Restriction items were similar in form to the Love items. Scoring of responses to Restriction items was interpretative.

The Family Cohesion items consisted in simple, multiple-choice questions about whether parents join with each other to visit and entertain, enjoy casual leisure activities in the home, and go out on family excursions.

In addition to including buffer items in the question schedules, we asked children some questions about the sociological characteristics of their families, including parental occupations, size of family unit, age of family members, and variety of family possessions.

The preliminary form of the question schedule was divided into two parts. It was found that, when the 74 items were arranged in a time sequence, they fell into a natural division of questions about weekday and weekend activities. We decided to use this natural division to provide a point at which administration could be broken.[5]

PRELIMINARY STUDY

The two preliminary forms were administered to a group of 45 fifth grade children in May 1956.

Administration

Because of the length and complexity of question schedules, it was found necessary to use experienced proctors. Proctors were used to ensure that children gave individual and meaningful answers to each question. Special features of the administration included reading each item aloud twice; emphasizing that children were to give the answer which described what *usually* happened; and presenting the schedule as a survey of the range of activities in which children of the local community engage. These and other aspects of the administration may be illustrated by the following excerpts from the standardized instructions given to the children before the Ordinary Day question schedule. "We are trying to find out the ways in which children spend their days. . . . The questions start with your getting up in the morning and end with your going to bed at night. We want you to give the answer which is *usually true for you.*"

Administration of the preliminary forms took two and a half hours. In addition, a 20-minute play break was given between administration of the

---

[5] The two question schedules were called "My Ordinary Day" and "My Usual Weekend."

Ordinary Day and Usual Weekend. The children enjoyed the idea of telling about their everyday activities and seemed to find most questions interesting. Form and phrasing of questions was on the whole appropriate: children gave satisfactory answers to all but a few items. Further, none of the children seemed disturbed by the content of the questions.

Scoring

Criteria for scoring the Love, Restriction, and Family Cohesion items were derived from a priori definitions of these variables. For example, criteria for scoring a Love item about the quality of parental reactions to especially good child behaviors were that parental interest be shown by the giving of praise; that interest be shown by mother *or* father or *both*;[6] and that praise be expressed by giving appropriate rewards. Similarly, in scoring a Restriction item about whether children are expected to confine their activities to specified locations, we first established that the answer indicated the existence of a restriction, but then ascertained whether the restriction was a defensible one. For example, we did not score an item as an indicator of Restriction if places like swiftly running creeks at which many fatalities occur were forbidden to children. Since Family Cohesion items were in multiple choice form, scoring was largely mechanical.

Findings

An internal consistency analysis[7] showed that 40 of the 74 items could be arranged into four clusters.

Love. It was found that only three of the eight items designed to measure this variable were strongly related to each other.

Restriction. The items intended to measure this variable divided into two clusters. The first cluster consisted in six questions about the extent to which children are permitted to undertake activities and social relationships of their own choosing and to choose the venue of some of these. We decided to call this cluster of items *Social Restriction*. The remaining nine questions were about whether children assist their parents in routine household activities and/or take responsibility for these. This cluster was named *Household Responsibilities and Duties*. Since these two item clusters were unrelated, it seemed that our a priori assumption that excessive participation in household routines must restrict children is probably untenable.

Family Cohesion. This consisted of a cluster of 22 items. It was found that in more than 80 per cent of cases in which parents are reported to share activities, children state that they, too, participate in these activities.

In all cases, the clusters of items which comprised the various measures of our variables were found to be independent of each other and of all other

---

[6] Initially, no attempt was made to discriminate between the behavior of fathers and mothers.

[7] This analysis was undertaken on the Hammond Matrix Sorter.

items in the question schedules. All items comprising a cluster were significantly correlated with the total cluster score.

SECOND STUDY

The findings from the preliminary study encouraged us to continue our attempts to describe selected aspects of the parent-child relationship in terms of children's reports of their daily lives with their parents. In the second study, we aimed, first, to cross-validate the findings concerning the independence and internal consistency of the clusters of items representing the Social Restriction, Household Responsibilities, and Family Cohesion variables on a larger and more representative sample of children. Further, we wanted to continue the attempt to construct a set of items to measure Love.

The Ordinary Day and Usual Weekend schedules were reduced in size by rejection of nondiscriminating items. Sociological and buffer questions were retained. Five of the eight items designed to measure Love were reworded, and scoring criteria sharpened. In all, the revised form of the question schedules contained 59 questions, 45 of which were directed at the measurement of Love, Social Restriction, Household Responsibilities and Duties, and Family Cohesion variables.

The revised schedules were administered to a second sample of 150 fifth grade boys in July 1956. This sample was drawn from four schools, two of which were in residential suburbs, while the other two were in or adjacent to industrial areas. The administrative procedures were unchanged, but the time taken to administer both revised question schedules was reduced to an hour and a half.

An analysis of the eight items intended to measure Love showed that six of these formed a cohesive cluster. This cluster consisted of the three items which were found to be satisfactory in the preliminary study and three of the five revised items. A further examination of the six items comprising this cluster showed that the average intercorrelation of the items was of the order of .81. This very high degree of internal consistency was obtained from items asking children about the extent to which their parents' behavior was oriented to their needs, respecting of their personality characteristics, and designed to guide them in their relationships with their environment.

In analyzing the relationships between the four item clusters, the sample of 150 boys was divided randomly into two groups. Since we had examined intact school grades, we could not assert that this sample was typical of children of the age group being studied: for instance, it was possible that one of the schools examined had an undue proportion of children from a homogeneous socioeconomic background. In both samples, it was found that the item clusters which constituted the measures of the four main variables were independent of each other.

TABLE 1

CORRELATION COEFFICIENTS BETWEEN FAMILY VARIABLES

	Love	*Social Restriction*	*Household Duties*	*Family Cohesion*
Sample 1 ($N=75$)				
Love		—.17	.06	.01
Social Restriction			—.06	—.11
Household Duties				—.01
Family Cohesion				
Sample 2 ($N=74$)				
Love		.07	.02	—.06
Social Restriction			.08	—.12
Household Duties				.05
Family Cohesion				

NOTE.—The samples are unequal because one case was rejected.

The findings summarized in Table 1 replicate the independence of the Social Restriction, Household Responsibilities and Duties, and Family Cohesion dimensions. This table also shows that the Love item cluster is independent of the other three.

The earlier finding that, within a given cluster, all items were related to the total score at at least the .05 level of significance was replicated in this second study. In addition, it was found that the scoring was capable of being reproduced by an individual unfamiliar with the question schedules. The obtained index of interscorer reliability was 97 per cent.

THIRD STUDY

In this study we planned to investigate some correlates of scores on the four family scales if the finding that the scales were independent of each other was cross-validated.

The third sample was selected to meet the same general specifications as those set out for earlier samples: namely, only literate 10- to 11-year-old boys, living in intact families in the local urban community would be included. In this study, we made a serious attempt to ensure that the sample drawn should be large and that children from different sections of the community should be represented. However, it was impractical to examine members of one large parochial group. The sample may be considered to be representative of all other sections of the local community. In all, 519 boys attending 16 different schools were examined in February and March 1957. The schools were located in a wide variety of ecological areas: at least two schools were drawn from each of inner industrial, inner residential, outer industrial, outer residential and mixed areas.[8] Children from both government and nongovernment schools were included in the sample.

A number of measures of potential correlates of family variables were administered to the third sample. Some of the measures of the correlates were administered at the same time as the family question schedules, while the remaining correlate measures were administered two to three months later. For the purposes of investigating relations between the family scales and the correlate measures, we specified that, for a case to be included in the analysis, data had to be complete. If even one correlate measure was incomplete or invalid, the case was dropped. Unfortunately, a severe epidemic of Asian "flu" started in March 1957. This led both to delay in the administration of the second set of measures and to an extraordinarily high rate of absence from school, and the latter factor led to a large number of incomplete cases. In all, 432 complete case records were available for analysis.

TABLE 2

CORRELATION COEFFICIENTS BETWEEN FAMILY VARIABLES ($N = 432$)

	Love	*Social Restriction*	*Household Duties*	*Family Cohesion*
Love		—.08	.09	.03
Social Restriction			—.08	—.13
Household Duties				.12
Family Cohesion				

Analysis of the characteristics of the family scales showed that, for the third sample, the Love, Social Restriction, Household Duties, and Family Cohesion scales were relatively independent of each other. Relevant correlation coefficients are presented in Table 2.

Reliability of the Scales

Because of the relatively interpretative nature of the scoring, we rechecked interscorer reliability. Twenty-five cases were drawn at random from the original pool of 519 cases. Two persons who had not previously worked on records independently scored all the items constituting the family scales in these 25 records. The agreement between the two independent scorers was 99 per cent; moreover, each of these scorers registered 99 per cent agreement with the original scorers.

An opportunity arose to adminster the question schedules to 46 children who had participated in the second study. The time interval was eight

---

[8] The local council of the community in which the study was undertaken lays down regulations concerning the type and extent of industrialization permitted in a given area.

TABLE 3

RETEST RELIABILITY COEFFICIENTS FOR FAMILY SCALES ($N = 46$)

Scale	r
Household Duties	.78
Social Restriction	.71
Family Cohesion	.60
Love	.80

months. As the results in Table 3 show, all of the reliability coefficients are significant beyond the .01 level.

In addition to calculating these simple retest reliability coefficients, we examined the retest consistency of the single items. On the Household Duties scale each single item was consistent beyond the .001 level, despite the eight-month interval. Five out of the six items comprising the Love scale were consistent beyond the .001 level. While the remaining Love item was not significantly consistent, scores tended to indicate that the item was not wholly unreliable. In the case of Social Restriction, all six items were consistent beyond the .001 level. Only six out of the 12 items[9] constituting the Family Cohesion scale were significantly consistent.

Validity

While the data already presented show that it is possible to use children's statements to evolve measures of defined aspects of the parent-child relationship which are independent, internally consistent, and reasonably stable, no evidence has yet been presented to indicate that the measures are related to independent assessments of child or parent behavior.

A number of techniques for assessing child behavior and personality were administered to children comprising the sample used in the third study. An indication of the predictive utility of the family scales can be given by setting out some relationships found to obtain between the family scales and some of the independent measures of child behavior. A significant and meaningful relationship, $r = -.39$, between the Love scale and Reputed Immaturity (4) was obtained for a sample of 250 used in the third study. No other family scale was related to this sociometric measure. The Household Duties and Responsibilities scale was found to bear a linear relation with the external criterion of school marks. For a sample of 96 boys, drawn from four schools where marks had a comparable mean and standard deviation, $\chi^2 = 6.47$ with 2 df $(.05 > p > .02)$. School marks were not related to scores on any other family scale. Neither the Family

---

[9] The Family Cohesion scale was reduced from 22 to 12 items after the second study.

Cohesion nor the Social Restriction scales have so far shown any consistent relations with independent measures of child behavior.

It was also possible to relate scores on the family scales to data on parent and child behavior obtained during interviews with parents. Mothers of 69 of the children were interviewed by an experienced psychologist or social worker who knew no details of the work already done with the children. There were eight interviewers, no one of whom did less than six interviews. Although interviewers used their own chosen techniques of phrasing and ordering questions, they were provided with a schedule of topics. The interviews covered a wide range of topics including socioeconomic status of family, childrearing practices and beliefs, and parents' aspirations and plans for their children. Topics were included which gave mothers an opportunity to discuss their behavior and the rationale for their behavior in situations we thought sensitive to the expression of "Love" and "Household Duties and Responsibilities" variables. Mothers were also questioned about their children's actions and attitudes in these situations. The interviewers, who were not acquainted with the children's scores on the family scales, were asked to record interview data in a way that would allow the mothers' responses to be clearly differentiated from interviewer interpretations. The present writers jointly scored the mothers' recorded statements in terms of the a priori definitions of Love and Household Duties and Responsibilities. The scoring was blind, all identifying particulars having been removed from the interview records beforehand.

The interviews took place between 18 months and two years after administration of the family scales to the children. Nonetheless, on both the Love and the Household Duties and Responsibilities variables, scores from interview records were significantly related to scores from question schedules administered to children. On Love the correlation coefficient was .29, and on Household Duties and Responsibilities the correlation coefficient was .39.[10] When records obtained by different interviewers are examined separately, the relation between interview scores and child question schedule scores is seen to vary considerably: for instance, the correlation coefficients for Love range from .85 to —.09.

In view of these findings, it seems reasonable to conclude that the two scales measuring Love and Household Duties and Responsibilities are psychologically meaningful as well as internally consistent and independent of each other. Despite the internal consistency and independence of the Social Restriction and Family Cohesion scales, we cannot present unequivocal evidence of the validity of these scales.

More generally, the results presented in this article indicate that, if care is taken in the selection and wording of questions, 10- to 11-year-old children can give reliable and valid information about defined aspects of the parent-child relationship. While no claim is made for universal validity of the variables isolated and measured in this study, results obtained do seem to encourage the view that construction of measuring devices can profitably follow from a priori definition of variables.

---

[10] For the sample of 69 cases, the Love correlation coefficient is significant beyond the .05 level, and the Household Duties and Responsibilities beyond the .01 level.

REFERENCES

1. BALDWIN, A. L. The effect of home environment on nursery school behavior. *Child Develpm.*, 1949, 20, 49-62.
2. BALDWIN, A. L., KALHORN, J., & BREESE, F. H. Patterns of parent behavior. *Psychol. Monogr.*, 1945, 58, No. 268.
3. BALINT, M. *Problems of human pleasure and behaviour.* London: Hogarth, 1956.
4. BESWICK, D. G., & COX, F. N. Reputed aggression and dependence in children. *Austral. J. Psychol.*, 1958, 10, 144-150.
5. CHILD, I. L. Socialization. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. 2. Addison-Wesley, 1954. Pp. 655-692.
6. ERIKSON, E. H. *Childhood and society.* Norton, 1950.
7. FREUD, A. *The ego and mechanisms of defence.* London: Hogarth, 1937.
8. FREUD, A., & BURLINGHAM, D. *Infants without families.* London: Allen & Unwin, 1943.
9. FRIEDLANDER, K. *The psycho-analytical approach to juvenile delinquency.* London: Routledge & Kegan Paul, 1947.
10. HERBST, P. G. Analysis and measurement of a situation: the child in the family. *Hum. Rel.*, 1953, 6, 113-140.
11. JERSILD, A. *Child psychology.* (4th Ed.) Prentice-Hall, 1954.
12. LEVY, D. M. *Maternal overprotection.* Columbia Univer. Press, 1943.
13. LEWIN, K. *A dynamic theory of personality.* McGraw-Hill, 1935.
14. LEWIN, K., LIPPITT, R., & WHITE, R. Pattern of aggressive behavior in experimentally-created "social climates." *J. soc. Psychol.*, 1939, 10, 271-299.
15. REDL, F., & WINEMAN, D. *Children who hate.* Free Press, 1951.

ADAPTABILITY: A PROBLEM SOLVING APPROACH TO ASSESSING CHILDREARING PRACTICES[1]

IRVING TALLMAN*
San Francisco State College

Theories concerning the consequences of parental practices on children's behavior have only recently been interpreted in terms of empirical observations. Present research in this field can be broadly categorized under three theoretical orientations: those concerned with specific parental behavioral dimensions which serve as antecedents to consequent personality and behavioral components in children (3, 20, 24); those which assess parental attitudes and the consequences of such attitudes (19); and, finally, those which attempt to assess interaction patterns and the differentiation of roles within the family (14, 17, 21).

This paper will report on the foundational work and research findings based on still another theoretical approach. In this approach emphasis is placed on the problem solving character of the childrearing process. The data to be reported are based on partial results of a study presently in progress at San Francisco State College. One phase of this study is concerned with the relationship of parents to their retarded children. What follows is not an advocacy of the superiority of a particular approach over other means of assessing parent-child relationships and practices. Rather, the purpose of this paper is to present a unique approach and one which harbors the potential for contributing an additional perspective to the understanding of the parent-child relationship.

A PROBLEM SOLVING FRAMEWORK

The conceptual framework to be presented has been developed principally out of the theoretical formulations of John Dewey, H. G. Mead, and W. I. Thomas. The basic assumption made is that the parent-child relationship exists as a series of problem solving situations. Such situations involve

---

* San Francisco State College, 1600 Holloway Avenue, San Francisco 27, California.

[1] The conceptual and empirical work reported in this paper was made possible by a grant from the United States Office of Education, Department of Health, Education, and Welfare, Contract No. SAE 6439. The writer wishes to acknowledge his indebtedness to Samuel Levine for his careful reading and criticism of the manuscript, as well as for his helpful comments and suggestions over the three years in which the ideas to be presented in this paper were formulated.

Child Develpm., 1961, 32, 651-668.

mutual expectations as to the behavior of each participant. From the point of view of the parent, he not only has expectations for the child in a situation but also of himself in terms of his parental role. His actions are seen as affecting or training the child to behave according to certain standards. A parent's actions can be conceived as having two purposes: the resolution of immediate problems and the socialization of the child.

A parent responding to a child in any situation would theoretically have an infinite number of behavioral acts to draw upon. That this is not the case is due to a number of factors which can best be summarized under the dual categories of personal experience and social structure. On the basis of past experience the parent has accepted some behavioral forms and discarded others. His life experiences have positively reinforced some forms of action and negatively reinforced others. The social structure, by means of normative standards, also delimits the choice of acts available. The normative pattern prescribes the rights and duties of a particular social position, and in so doing, structures the forms of interaction possible in any situation.

Within the limits of these experiential and social restrictions on behavior, we must assume that a wide range of choices of action still remain. No society could long exist if it did not allow sufficient behavioral flexibility to cope with constant changes and unexpected situations. From a problem solving point of view, the normative patterns can provide only the skeletal structure for effectively dealing with a situation. It follows that no *specific action* or form of behavior can be considered as right or best for any individual: first of all, there exists a range of possible "right" responses; secondly, the responses available to the individual are limited by experience and the normative patterns of the culture and the subculture. For example, a middle class mother may respond to a child's misbehavior by giving him "love" or reassurance, whereas a working class mother may spank her child under the same circumstances. Either action, if adhered to rigidly and mechanically might be fruitless and unsuccessful whereas, if it were appropriate to the situation, it could be effective. To consider the former as loving behavior and the latter as punitive would overlook the more crucial meaning of the behavior, e.g., its appropriateness to the situation.

The problem of raising a child is a day-to-day and moment-to-moment affair, and, within certain structural limits, it occurs in constantly changing and shifting situations. The action a parent chooses, therefore, is determined by his perception of the specific situation. It is the "definition of the situation" which W. I. Thomas has referred to as the essential stimulus for action (22). Interactions and transactions which occur in a situation effect constant changes and call for constant situational redefinitions.

This dynamic interpretation of behavior presents a conceptual problem to the researcher who must make his assessment at a particular point in time. The theoretical framework presented in this paper suggests at least a partial solution to the problem, that is, the recapitulation of the problem solving process. It is assumed that in facing any situation an individual will bring into play his experiences with similar situations and his awareness of alternative forms of action. These two factors in combination with the person's assessment of the situation will result in a plan and finally specific action. This process, hereafter referred to as adaptability, will be amenable to empirical exploration to the extent that it can be verbalized and observed.

The following pages will be concerned with explicating the concept of adaptability, defining its components, providing a set of operational definitions, and, finally, reporting data relevant to the utility of the concept to research in childrearing.

THE CONCEPT OF ADAPTABILITY

Baldwin (2) states that the term ability has been defined in two senses: it reflects what an individual "can" do, and it can also be used as a primary factor basic to a specific act. In either case it is considered as ". . . a necessary but not sufficient condition for actual performance" (p. 199). A distinction is therefore made between being able to accomplish a feat and actually doing so. Baldwin further states that ". . . ability is a meaningful concept only because it describes the characteristics of a person which contribute to his carrying out adaptive behavior" (2, p. 210). Adaptive behavior is "result oriented"; it is ". . . behavior oriented to the consequences of action." Thus, a problem solving analysis demands more than a specification of ability; it must include basic conditions under which adaptive behavior occurs. In childrearing situations we are interested in the appropriate abilities, *plus* other factors, which contribute to the parent's behaving *adaptively* toward his child. Baldwin suggests that one of the added requirements necessary to convert the ability into adaptive action is motivation. The writer would add to this the concept of environmental opportunity or environmental difficulty, for, as Baldwin recognizes, some situations may demand greater skill than others for achieving a similar result.

Four components are conceived of as essential to effective adaptive behavior toward a child. Two of these may be considered as abilities, the other two as factors which make it possible to employ the abilities effectively. These latter elements are motivation and the availability of resources in the environment. The two attributes which can be considered as abilities are empathy and flexibility. These four components combine to form what Burgess and Wallin (5) have titled "adaptability." The term aptly combines the two fundamental constructs which the writer considers as necessary for effective problem solving.

In order for behavior to be effective the four components of adaptability are conceived of as acting in combination. Motivation is seen as the basic prerequisite. If the parent does not wish to do anything about a situation, we assume that nothing will be done, whatever his potential problem solving skills. Conversely, the parent may be highly motivated, but if he does not have the understanding and flexibility necessary to determine the appropriate action and respond accordingly, the probability is that the action taken will not be adaptive. In addition, neither empathy nor flexibility should produce effective results without benefit of the other. The ability to understand the child and the situation (empathy), while lacking a repertory of alternative actions (flexibility), appears as fruitless as having a wide repertory of actions without such action being guided by appropriate understanding.

Only three of the four components are employed in the actual determination of adaptability. "Availability of resources" is difficult to assess and is, at least in part, incorporated in the construct of flexibility.

The research instrument developed, therefore, was designed to measure independently the components of motivation, empathy, and flexibility. In this way it was possible, indirectly, to test the hypothesis that these components are interdependent. The findings relevant to these interrelations will be discussed in a subsequent section of this paper.

The Problem of Definition

Terms such as motivation, empathy, and flexibility are encumbered with a great degree of ambiguity. The problem is to define each of the components with sufficient specification to meet the purposes of our particular conception of adaptability, while, at the same time, not destroying the constitutive meaning which does exist for these terms.

Parent adaptability is defined as the parent's ability to deal effectively with problematic situations by changing roles, attitudes, and actions in terms of new or modified understandings of the situation with which he is confronted.[2] With this definition in mind we will turn to an examination of each of the components.

Flexibility is defined in both attitudinal and behavioral terms. It represents an attitude which individualizes the child in the light of a specific situation, as opposed to stereotyping both the child's behavior and the situation. In addition, it represents the ability to employ alternative means of action toward one's child in the light of the parent's evaluation of the situation, as opposed to acting in a rigid, stereotyped manner.[3] This continuum involves more than the use of alternative modes of actions in dealing with a situation; it also includes the evaluation of these actions and the integration of attitudes with stated actions. The lower end of the flexibility continuum is characterized by absolutism and arbitrariness; the higher responses indicate an ability to evaluate relative and pertinent influences. To this extent the scheme approximates Rokeach's concept of the "open and closed mind" (16). In addition, low flexibility contains attributes which contribute to what Frenkel-Brunswick has termed "intolerance of ambiguity" (8). Therefore, a response rated as low flexibility would be one which stereotypes the child in the situation and employs stereotypic statements to rationalize the action or lack of action taken. Thus, a statement such as "retarded children need to be loved, so what we do is love him" would be rated low. A middle response would be a single behavioral response to the specific situation, for example, "If he interrupts me when I talk with my husband I'll say, 'I'm talking now, wait until I finish.'" A high response is one which chooses a specific behavior from a number of alternatives, the behavioral choice being dependent on the parent's specific evaluation of the situation, for example, "I find if you show him how to do it by taking his hand and doing it with him, from then on he knows what to do. It's much easier than telling him because it's hard for him to understand."

---

[2] This definition of adaptability is a modification of that presented by Burgess and Wallin (5). These writers define adaptability as "the capacity of the person to change his roles, his attitudes, and his behavior in order to adjust to those of other persons or to new or modified situations" (p. 623).

[3] The definitions of six levels on the continua for each of the components of Flexibility, Empathy, and Motivation and the interview schedule have been deposited as Document number 6838 with the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C. A copy may be secured by citing the Document number and by remitting $2.50 for photoprints, or $1.75 for 35 mm. microfilm. Advance payment is required. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress.

Empathy has variously been defined as "a gift of understanding people" (1), "social sensitivity" (10), and "the ability to take the role of the other," an ability which Sarbin (18) subsumes under what he calls an "as if" skill. Our interpretation of empathy adheres to these general definitions. It differs, however, from the usage given to it by Dymond (7), Notcutt and Sylva (14), and others in that we make no stipulation as to the accuracy of perception. An attempt to assess perceptual accuracy would involve delimiting the situational cues a parent may have to just those areas in which the investigator has factual information. Such a procedure would be a digression from a problem solving orientation. In addition, as Bruner and Taguiri (4) have pointed out, perception of emotions requires as many cues as possible, as well as an understanding of the situational context in which the perceived behavior occurs. Our interest is in the depth and inclusiveness of the parent's perception of the child, within the context of a situation. The assumption is that inclusiveness and the integration of pertinent situational factors in understanding the child is sufficient evidence of sensitivity to infer the ability to take the child's role and consequently to justify the use of the term empathy.

Empathy is defined, for our purposes, as the parent's ability to interpret his child's behavior in terms of feelings, thoughts, and motives and to relate this understanding to the specific situation in which the parent and the child are involved. The lowest point of the continuum represents the inability of the parent to think or express why the child behaves as he does in a given situation. The midpoint is indicated by the parent's making a single or simple cause-and-effect relationship between the child's activity or response and the stimulus which elicited the response; the highest level is understanding the child's feelings, thoughts, and motivations and combining this understanding with the stimulus situation in order to explain the behavior.

The term motivation has perhaps the widest range of interpretation and the least amount of constitutive meaning of any of the components. Its breadth and complexity is dramatically summarized by Murphy's statement that "Any serious conception of social motivation must be theoretically complex enough to provide raw material for the manifold subtle components in the motive potential of mankind, must see how the growth and learning process may complicate and interweave these expressions, and how the different types of cultural arrangements may mold, stabilize, and integrate the various fusions and complications of individual experience" (12, p. 613).

The magnitude of the task, as outlined by Murphy, is overwhelming and goes beyond the pragmatic requirements for assessing a component in adaptability. The statement illustrates the theoretical and communicative problems resulting from employing the idea of motivation. What is necessary is to specify the *kind* of motivation that is of interest. We are not concerned with basic or genetic drives; our focus lies in a particular form of derived social motive, specifically the motivation of the parent to solve interpersonal problems in which he and his child are interacting. A consequence of such motivation is involvement. Involvement represents the degree to which the parent gives of his time and energy in order to achieve certain results. Motivation, for parent adaptability purposes, is therefore defined as an affectual investment characterized by involvement of the parent in effecting certain anticipated goals for the child. "Affectual investment" is inferred from the behavioral interpretation of involvement. Involvement is conceived as a continuum, its lowest level being an unwillingness on the part of the parent to recognize any responsibility for the child or the situation, i.e., "There is nothing you can do with him." At the other end of the continuum, high involvement includes evidences of increased effort, such as inclusion of external resources and engaging in persistent follow-through by means of interaction with the child, i.e., "I went to school and watched the teacher work with him. She's calm, but she's firm and she expects him to sit quietly and eat, so I expect better behavior from him at the dinner table. If he doesn't behave, he must leave the table. He'll object, but I insist." The middle of the continuum is represented by a willingness on the part of the parent to stay with the situation over a period of time, i.e., "I always talk to her and warn her against going out in the street."

Underlying the concept of adaptability, and consequently each of the components, are two continua. The first conceives of the parent's perception of the situation running from generality to specificity. In line with the conception of behavior as being determined by the definition of the situation, the assumption is that, the greater the ability to specify the elements in the specific situation, the greater the potential for problem solving.

TABLE I

RELATION OF THE GENERALITY-SPECIFICITY AND COGNITION CONTINUA TO EACH OF THE ADAPTABILITY COMPONENTS

	ADAPTABILITY COMPONENTS			UNDERLYING CONTINUA FOR ALL COMPONENTS	
	Motivation	*Empathy*	*Flexibility*	*Generality-Specificity*	*Cognition**
Low	No acknowledgement of responsibility for personal involvement.	No recognition or thoughts about causation of child's behavior.	Rigid and stereotyped attitudes and actions.	Categorical or general response.	"Unguided behavior."
Middle	Staying with a situation by a particular type of action.	Making a simple and single connection between cause and effect of children's behavior.	Use of a specific or limited act or applied formula.	Limited specificity.	"Simple guided behavior."
High	Integration of parents' efforts with outside resources; interpersonal interaction with child with persistent follow-through.	Understanding of the child's feelings, motivations, and thoughts as well as the situational factors which resulted in behavior.	Dealing with the situation in terms of its specific meaning, while employing past experience and recognizing changes in the child.	High specificity.	"Cognitive guidance."

---

* Based on Baldwin (2).

The other underlying continuum evolves from Baldwin's concept of adaptive behavior as behavior oriented to the consequences of action (2). If this is the case, then the more the behavior is what Baldwin has termed "cognitively guided," the greater is the probability of effective problem solving. The continuum is conceived as moving from behavior which would remain unvaried under any circumstances (i.e., unguided); through a mid-point in which the behavior is guided (i.e., a simple reaction to the situational stimuli); to behavior which is cognitively guided (i.e., consciously employed to effect specific and long range goals).

As Table 1 indicates, these underlying continua are common to the components of flexibility, empathy, and motivation. They are a priori determinations which are integral to the concept of adaptability.

Operational Interpretation and Instrument Development

It can be seen, from the above discussion, that the conceptual framework of adaptability places great emphasis on the conscious thought processes of the parent. The assumption is that people who can think in adaptable terms are more inclined to act in a problem solving manner. If we make the additional assumption that people who are able to express themselves in adaptability terms are more likely to act in this way, we can think of employing an interview schedule as our research instrument. The test of this assumption lies in the experimental use of the instrument. Empirical support of the postulates to be tested can be considered as indirect support of the instrument and its assumptions. However, if the postulates are not supported, what are we to say about the assumptions and the instrument? Will the results be due to an inappropriate instrument, or have our theoretical constructs been disproved? The writer, at this point, cannot argue for the proven utility of the research instrument. He can, however, point out that in his use of such an instrument the results supported the hypothesis that the children of highly adaptable parents showed greater growth on a measure of social competence than children of lower adaptable parents.

An interview was devised to elicit parental responses on each of the three adaptability components. The questions were open ended and worded in a simple and direct manner. The respondent was presented with a problematic parent-child situation, such as behavior at the dinner table, and then asked specific questions designed to educe flexibility, empathy, and motivation responses.[4] Every attempt was made to make the interview as natural as possible while at the same time maintaining a standardized and structured procedure. The interviewer introduced a situation as follows: "I'd like to talk with you now about Johnny's behavior at the dinner table." After asking a general question about the parent's impression of Johnny's table manners, the flexibility question was asked in this way: "How do you go about training Johnny in table manners?" If necessary, it was followed by a further probe, "Is this what you always do?" Then, the empathy response was elicited by asking, "Why do you think Johnny behaves (as parent has previously stated) at the dinner table?" The motivation questions were somewhat more complex. First, an attempt was made to obtain the parent's level of aspiration for the child in the given situation, e.g., "What would you hope for (or like) in regard to Johnny's behavior at the dinner table?" The parent was then asked two questions designed to determine the amount of involvement evident in achieving these goals for the child. The next question was, "What do you think can be done to help achieve this?" and, if necessary, "What have you done in this regard?"

---

[4] *See* Footnote 3.

This, in general, was the procedure employed with the parents. Interviewer bias was, of course, one of the principal sources of error in this type of approach. In an attempt to control as much of this error as possible, interviewers were given fairly extensive training in the use of the schedule. The schedule was adhered to rather rigidly, and interviewers were allowed to probe only under specified conditions. Interviewers were all professional social workers, clinical psychologists, or educational psychologists.

In the present study verbatim transcriptions were made from the tape recorded interviews, and the subject's responses for every question were entered on separate slips of paper. The slips were stamped with a random number representing the subject. Responses were then grouped according to the question number and component. Judges would rate 50 such responses at a time. Judges were paired, each making an independent rating for every response and then meeting with his partner to discuss and resolve any disagreements. No pair of judges rated all responses for any group of subjects. This method would seem virtually to eliminate halo effect. In all, judges rated 10,000 separate responses. The reliability for independent ratings of pairs of judges is high (correlations were consistently above .90). Judges rated responses according to the criteria established for each of the components along 6-point continua.

In addition to the above procedure, judges made ratings on the basis of listening to the tape recorded interviews for a random sample of 30 parents. The correlations between the two rating procedures was .84. This suggests that the rather cumbersome and expensive method employed to avoid halo effect may be unnecessary.

Scoring Procedures

The method for obtaining an individual's empathy, flexibility, or motivation score derives from the procedure employed in developing the instruments and, in particular, the method of selecting the questions. This selection procedure was, in turn, related to certain assumptions relevant to the conception of adaptability.

Questions were selected on the basis of 60 pre-experimental interviews. An original pool of 43 situations were compiled, reflecting the opinions of staff members as to common yet problematic areas of parent-retarded child interaction. In addition, the literature was reviewed, and other scales and schedules were examined for further cues as to relevant situations.

Five revisions of the interview schedule were made, based on the 60 pre-experimental interviews. The final schedule included 24 different situations.

There were three principal criteria for the inclusion of a question: (a) the degree to which the question elicited responses pertinent to our constructs of flexibility, empathy, and motivation; (b) the frequency with which parents indicated the situation occurred; (c) the range of responses the question pulled which differentiated among respondents. It will be noted that no attempt was made to order the questions; they merely represent a sample of situations found to be relevant to parents of retarded children. The response to *each* question is scored on the basis of an a priori ordered 6-point scale, based on the theoretical constructs relevant to the particular component being measured. However, inasmuch as the questions themselves cannot be ordered as to meaningfulness, difficulty, intensity, etc., the procedure for arriving at an individual score must be classificatory (23). In other words, this procedure does not allow us to meet the axioms necessary for an ordinal or interval scale; and, therefore, if we were to sum the questions scores in order to obtain a component score, we would be doing violence to the logic of measurement.

The decision to use the above procedure of item selection, rather than developing a formal scale along the lines of Guttman or Thurstone, stems from the theoretical assumptions regarding the nature of abilities. We assume that individuals seek to resolve problems as quickly and economically as possible. Therefore, even the most adaptable individual will respond only to the degree necessary to achieve his ends. Thus, if a situation does not represent a serious problem, the parent will not be affectually aroused and therefore not overly involved. As a consequence he will probably respond in a routine manner. The same circumstance will hold true if the situation has been resolved in the past and no longer presents a problem.

It is only on the questions which pose situations that are perceived as problematic by the parent that we hope to differentiate between low and high adaptable parents. The high parent, it is predicted, will respond to the challenge by an increase in cognitively guided problem solving activities; the low adaptable person will respond by increased rigidity and stereotyping. The former will react progressively; the latter will respond regressively.

.It can be seen from the above discussion that there is no means for predicting which situational questions will be perceived as problems for any respondent. What is claimed is that the questions, having been adequately pretested, represent a sufficient sample of common problematic situations so that a proportion of them will be pertinent to almost every parent.

The scoring procedure assumes that at least a majority of the questions will evoke problematic experiences in any parent. If there are 15 questions for a component, a highly adaptable parent should score high on at least five or more of the questions for that component. Conversely, a low adaptable parent will score low on five or more of the questions. The probability is that the parent who is given an over-all high rating for a component will score no lower than 3 (middle) on a given question; nor will a parent who received a low component rating score higher than 3 on a given question. The scoring procedure may be summarized as follows: (a) Judges rate parent's responses to each question on a 6-point scale. (b) Parent's ratings are then examined and the frequencies of 1 and 2, 3, 4, and 5 and 6 responses computed. If the subject has five or more questions scored at the 1 and 2 levels, he is ranked low; if five or more ratings are at the 5 and 6 levels, he is ranked high. If the subject has five or more level 4 responses he is given an upper middle ranking. All other distributions are given a middle ranking.

The validity or utility of this categorical system must rest on its ability to predict to an external criterion. Unfortunately, there is no exact method of determining "internal validity."[5]

Scores for each of the three components are arbitrarily combined by summation. This procedure is employed principally for the sake of convenience and in lieu of more precise knowledge of how these factors combine. However, there does appear to be sufficient evidence to support combining the components in some form. Table 2 presents the correlations for each of the questions with the components and with total adaptability.[6]

It appears that each of the items contributes sufficiently, not only to its component, but also to total adaptability. There are, however, a few questions which do not appear to be contributing either to component scores or to adaptability (such questions would, of course, be eliminated in any subsequent use of the instrument). In addition, the instrument seems more consistent for mothers than for fathers.

Table 3 reports the correlation of each of the components with total adaptability. Again, it can be seen that, although each component is operating independently, there is a significant association with total adaptability. In addition, the three components produce significant associations with an external criterion only when they are combined.

---

[5] Measures of internal consistency such as odd-even reliability and item total score correlations (*see* Table 2) were sufficiently high to support the conventional claim for a homogeneous scale. (The odd-even Pearson correlation, using the Spearman-Brown correction for attenuation, was .84 for mothers and .81 for fathers.) As noted above, the writer does not claim the adaptability instrument to be a scale in the technical sense; these correlations do suggest, however, that the situational questions are appropriate and meaningful to most parents.

[6] All correlations reported are gamma coefficients, unless stated otherwise. Gamma is a nonparametric statistic which measures the directed order of association, the degree to which the order of x is generally the same as the order of y. Specifically:

$$\text{Gamma} = \frac{\text{probability of like orders minus the probability of unlike orders}}{\text{sum of like and unlike probabilities}}$$

TABLE 2

ITEM CORRELATIONS (GAMMA) WITH SUBSCALES AND TOTAL SCORE FOR MOTHER AND FATHER SAMPLES

	MOTHER ($N=113$)				FATHER ($N=100$)			
*Item**	*Flex.*	*Emp.*	*Mot.*	*Adapt.*	*Flex.*	*Emp.*	*Mot.*	*Adapt.*
Flexibility Items								
2.	.53	.42	.36	.51	.53	.27	.21	.42
3.	.52	.37	.40	.51	.41	.48	.26	.30
4.	.46	.29	.28	.41	.55	.26	.29	.42
5.	.37	.24	.25	.33	.19	.03	.09	.09
7.	.58	.26	.34	.45	.55	.29	.19	.38
8.	.59	.32	.38	.50	.50	.34	.21	.47
9.	.36	.24	.20	.30	.48	.21	.14	.32
10.	.63	.45	.38	.54	.48	.25	.22	.36
11.	.48	.38	.37	.48	.37	.09	.25	.24
13.	.55	.32	.31	.45	.52	.27	.35	.46
16.	.61	.31	.40	.50	.25	.16	.16	.22
17.	.38	.27	.13	.30	.40	.05	.13	.22
18.	.41	.23	.11	.31	.30	.20	.39	.34
19.	.27	.04	.08	.16	.33	.23	.30	.33
20.	.42	.34	.34	.42	.48	.06	.19	.26
Empathy Items								
2.	.32	.15	.21	.27	.26	—.004	.20	.18
3.	.21	.18	.17	.21	.28	.27	.23	.32
4.	.26	.06	.16	.20	.32	.02	.08	.16
5.	.46	.22	.25	.35	.35	—.06	—.05	.00
7.	.34	.28	.29	.36	.39	.13	.18	.30
9.	.40	.19	.16	.29	.46	.21	.27	.40
10.	.34	.05	.31	.24	.28	—.05	.07	.16
11.	.48	.24	.38	.42	.39	.23	.04	.31
12.	.42	.39	.28	.42	.42	.23	.23	.39
13.	.46	.36	.40	.46	.36	.50	.20	.47
16.	.46	.28	.20	.35	.49	.36	.26	.52
17.	.47	.27	.23	.38	.40	.10	.02	.27
18.	.46	.35	.18	.37	.55	.37	.24	.54
19.	.43	.22	.34	.38	.56	.23	.24	.49
20.	.49	.29	.37	.45	.47	.30	.30	.47
Motivation Items								
6.	.31	.20	.32	.32	.45	.30	.18	.43
9.	.44	.30	.28	.39	.58	.25	.10	.42
11.	.50	.32	.34	.48	.42	.12	.14	.26
14.	.35	.22	.26	.31	.41	.25	.32	.41
15.	.37	.34	.30	.40	.34	.21	.03	.26
17.	.53	.16	.36	.39	.43	.26	.22	.39
20.	.35	.30	.24	.35	.49	.25	.25	.41
21.	.43	.32	.35	.44	.48	.30	.20	.47
22.	.48	.22	.20	.36	.43	.24	.31	.37
23.	.48	.33	.36	.44	.34	.25	.08	.29
24.	.44	.24	.25	.38	.33	.17	.18	.24
25.	.51	.25	.23	.40	.41	.45	.28	.37
26.	.16	.27	.42	.20	.37	.11	.23	.27

* Item refers to question on Parent Adaptability Schedule.

TABLE 3

ADAPTABILITY SUBSCALE INTERCORRELATIONS (GAMMA)

	Motivation	*Flexibility*	*Adaptability*
Mothers ($N=113$)			
Empathy	.48	.49	.72
Motivation	..	.48	.75
Flexibility	..	..	.75
Fathers ($N=100$)			
Empathy	.35	.31	.71
Motivation	..	.32	.70
Flexibility	..	..	.58

A MEASURE OF CONSEQUENCES OF ADAPTABILITY

The assessment of parent adaptability—or for that matter, any skill or ability—is of little value unless there is some means of evaluating the consequences of that ability.

One of the principal interests in the present research is the effect of parents on the growth and development of retarded children. The focus is on changes in children's social competency over a period of time. At present, we have data relative to changes over a one-year period; at the conclusion of the project we will have additional information covering a second year.

The instrument used for assessing such changes is the San Francisco Social Competency Scale.[7] This scale provides an objective check list by which a parent rates a child on a number of behaviors under four classifications: self-help, initiative-responsibility, social skills, and communication. The specific behaviors which make up the items of the scale were judged by groups of educators and psychologists, as well as parents, to be critical to the child's adequate functioning in the home situation. The reliability coefficients between mothers and fathers on independent ratings of the child were in the .80's. Reliability for mother ratings (fathers did not participate in this phase of the project) over a period of a year was .84. It seems, therefore, that this scale represents a sufficiently objective and stable measure of certain critical behaviors for retarded children.

---

[7] This scale was developed under the auspices of the U.S. Office of Education at San Francisco State College for the research project in mental retardation mentioned above. A discussion of its technical properties would prove lengthy and would be peripheral to the purposes of this paper. The interested reader may find this information in Leo Cain, Samuel Levine, *et al.*, "Study of the Effects of Special Day Training Classes for the Severely Retarded Child." Report to the Commissioner of Education, U.S. Department of Health, Education, and Welfare, June 1958 and June 1959.

Table 4 reports that highly adaptable mothers have children who indicate greater growth over a year's period on social competency than do middle adaptable mothers. Low adaptable mothers were eliminated from this analysis for reasons which will be discussed subsequently. Fathers were also eliminated because they did not participate in rating their children after one year.

TABLE 4

RELATION OF HIGH AND MIDDLE ADAPTABLE MOTHERS TO CHANGES IN CHILDREN'S SOCIAL COMPETENCY RATINGS OVER A ONE-YEAR PERIOD

Changes in Children's Social Competency Scores	HIGH ADAPTABLE MOTHERS		MIDDLE ADAPTABLE MOTHERS	
	N	%	*N*	%
High (+15 and above)	14	48	11	26
Middle (0 to +14)	10	35	13	30
Low (−1 and below)	5	17	19	44
Total	29	100	43	100
		$\chi^2 = 6.44$*		

* $p < .05$ for 2 *df*.

The principal conclusion that can be reached from the above table is that retarded children of highly adaptable mothers show greater growth on social competency than mothers ranked as middle adaptable. It must be remembered that we have no independent way of measuring adaptability; it merely represents the sum of the component scores for flexibility, empathy, and motivation. The proposition that adaptability consists of the combination of these three factors seems to be supported by the intercomponent correlations reported in Table 3.

The fact that the criterion measure for adaptability was associated with change in the social competence of severely retarded children tends to strengthen the generalization that parent adaptability is a factor in child development. Maturation in such children is slow, and changes over a year's time are small, compared to normal children of similar chronological ages (5 to 12). It is conceivable, therefore, that, given a relation between parent adaptability and growth in such children, this relation could be as great or greater with children whose growth potential is not so severely restricted.

Previous research has emphasized the role of social class variables in differentiating parents' childrearing practices (6, 11). These results suggest the possibility that the relation between mother adaptability and change scores in children's social competency is spurious, reflecting the association of both these variables with social class. Present results, however, indicate that mothers' adaptability is not highly correlated with social class variables. As indicated in Table 5, correlations of mother adaptability and family income, husband's occupational prestige, and mother education are not statistically significant. The distribution within each of these social class variables was adequate to yield a correlation if one had existed.

On the basis of these data, we can conclude that the relation between mother adaptability and change scores in children's social competency is independent of social class.

TABLE 5

MOTHER ADAPTABILITY CORRELATIONS (GAMMA) WITH SOCIAL CLASS VARIABLES

	Family Income ($N=77$)	*Husband's Occupational Prestige** ($N=78$)	*Mother's Education* ($N=77$)
Mother Adaptability . .	.22	—.07	.11

NOTE.—N's differ because of missing data.

* Derived from the breakdown of the N.O.R.C. data in Kahl (9).

In addition to the above relations, our findings shed light on some characteristics of low adaptable parents. It was found that such parents, as a group, tend to be less consistent in rating their children on social competency than do middle or highly adaptable parents. The ratings of low adaptable parents were associated with more extreme change scores in both directions over a year's period than were ratings of parents in the other categories. An F test was computed to determine the extent of the difference in variability of the social competency ratings between the lower and upper quartiles of mothers' adaptability scores. The F ratio was 2.65 in the predicted direction with 21 and 23 degrees of freedom, significant at the .05 level. As we interpret this finding, it appears that the low scoring parents, perhaps because of less involvement and less understanding of their children, were not able to rate accurately their children in terms of specific behavioral referents. The result was the greater variance and inconsistency in the ratings.

On the basis of these findings, we questioned the accuracy of the social competency ratings of low adaptable mothers. As Table 6 illustrates, a chi square computed between three classifications of children's change scores and all three groups of mothers not only reduces the significance in the difference between these groups, but also portrays the randomness of the low adaptability mother ratings. This finding in itself would be of little significance. However, when it is added to the meaningful differentiations which can be made by examining middle and high adaptable mothers' ratings, and when we consider the difference in variability in the ratings of low and high mothers, the conclusion of inaccuracy, inprecision, and in-

TABLE 6

RELATION OF MOTHERS' ADAPTABILITY TO CHANGES IN CHILDREN'S SOCIAL COMPETENCY RATINGS OVER A ONE-YEAR PERIOD

Changes in Children's Social Competency Scores	MOTHERS' ADAPTABILITY SCORES					
	HIGH		MIDDLE		LOW	
	N	%	*N*	%	*N*	%
High (+15 and above)	14	48	11	26	11	37
Middle (0 to +14)	10	35	13	30	10	33
Low (−1 and below)	5	17	19	44	8	27
Total	29	100	43	100	29	100
			$\chi^2 = 6.93^*$			

* $p < .20$ for 4 *df*.

consistency among low adaptable mothers appears to gain some support from these data.

SUMMARY

The writer in this paper has attempted to delineate some of the theoretical and methodological considerations underlying the development of an instrument designed to assess parents' problem solving abilities in childrearing situations. Such skills are conceived of as including, in combination, the attributes of empathy, flexibility, and motivation. This complex of components is considered to determine the individual's ability to deal effectively with problematic situations by changing roles, behaviors, and attitudes when required, a skill which has been labeled here as adaptability.

Results of the research reported here have shown a significant relation between parents' adaptability and growth in the social competency of retarded children. In addition, evidence is presented to support the contention that low adaptable parents are less precise, consistent, and accurate in their perception of their children's behavior.

Mother-father differences in adaptability have not been discussed in this paper, although there is evidence indicating that such differences exist. Analysis and additional information is being compiled on this subject. A discussion of adaptability as related to parental roles has been deleted principally because the possible explanations and generalizations are inclined to be limited to questions of rearing a retarded child, and therefore, not within the scope of this paper. The possibility exists, however, that questions of the nature of a parent's identification with his child may tend to specify the conditions under which a parent may be more or less adaptable.

The parent adaptability instrument attempts to assess childrearing within the parent's personal experiential and normative framework. This is done by considering the differential use of cognition and specificity with which the parent confronts and responds to a situation, rather than examining specific behaviors. This approach offers the possibility for re-evaluating the implications of social class differences in childrearing (6, 11). For example, the differences reported by Maccoby and Gibbs (11) for middle and working class parents in such areas as techniques of discipline may be reinterpreted from a problem solving point of view. The consequences of using different techniques of discipline may vary in relation to the expectations and meaningfulness of the act to the child and the parent. It may be as ineffective for a middle class parent to be physically punitive as for a working class parent to use withdrawal of love as a disciplinary measure. If we ignore the specific type of act employed, but, rather, approach the situation in terms of the appropriateness of the act based on the parent's understanding of the child and the situation and the parent's involvement in doing something about the situation, our inference may be different. As a matter of fact, present results indicate that, for mothers, adaptability is not highly correlated with social class variables.

The procedure for determining adaptability depends upon a classificatory system of measurement, which is lacking in precision and involves a considerable amount of error. Nevertheless, the present research indicates that it is capable of producing fruitful results in assessing childrearing practices. In addition, the concept of adaptability may have heuristic implications in assessing the parental contribution to behavioral disorders in children.

References

1. Allport, G. W. The historical background of modern social psychology. In G. Lindzey (Ed.), *Handbook of social psychology.* Addison-Wesley, 1954. Pp. 3-56.
2. Baldwin, A. L. Role of an ability construct. In D. C. McClelland (Ed.), *Talent and society.* Van Nostrand, 1958. Pp. 195-232.
3. Baldwin, A. L., Kalhorn, J., & Breese, F. H. Patterns of parent behavior. *Psychol. Monogr.,* 1945, 58 (Whole No. 268).
4. Bruner, J. S., & Tagiuri, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology.* Addison-Wesley, 1954. Pp. 634-654.
5. Burgess, E. W., & Wallin, P. *Engagement and marriage.* Lippincott, 1953.
6. Davis, A., & Havighurst, R. J. Social class and color differences in child rearing. *Amer. sociol. Rev.,* 1946, 11, 698-710.
7. Dymond, R. A scale for the measurement of empathic ability. *J. consult. Psychol.,* 1949, 13, 127-133.

8. Frenkel-Brunswick, E. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, 18, 108-143.

9. Kahl, J. A. *The American class structure.* Rinehart, 1957.

10. Lindzey, G., & Borgatta, E. F. Sociometric measurement. In G. Lindzey (Ed.), *Handbook of social psychology.* Addison-Wesley, 1954. Pp. 405-448.

11. Maccoby, E. E., & Gibbs, P. K. Methods of child-rearing in two social classes. In W. E. Martin & C. B. Stendler (Eds.), *Readings in child development.* Harcourt, Brace, 1954. Pp. 380-396.

12. Murphy, G. *Social motivation.* In G. Lindzey (Ed.), *Handbook of social psychology.* Addison-Wesley, 1954. Pp. 601-633.

13. Newcomb, T. *Social psychology.* Dryden, 1950.

14. Notcutt, B., & Silva, A. L. Knowledge of other people. *J. abnorm. soc. Psychol.*, 1951, 46, 30-37.

15. Parsons, T., & Bales, R. F. *Family, socialization and interaction process.* Free Press, 1955.

16. Rokeach, M. *The open and closed mind.* Basic Books, 1960.

17. Rommetveit, R. *Social norms and roles.* Univer. of Minnesota Press, 1958.

18. Sarbin, T. R. Role theory. In G. Lindzey (Ed.), *Handbook of social psychology.* Addison-Wesley, 1954. Pp. 223-258.

19. Schaefer, E. S., & Bell, R. Q. Development of a parental attitude research instrument. *Child Develpm.*, 1958, 29, 339-361.

20. Sears, R., Maccoby, E. E., & Levin, H. *Patterns of child rearing.* Row, Peterson, 1957.

21. Strodtbeck, F. L. Family interactions, values, achievement. In D. C. McClelland (Ed.), *Talent and society.* Van Nostrand, 1958. Pp. 135-194.

22. Thomas, W. I., & Thomas, D. *Child in America.* Knopf, 1928.

23. Torgerson, W. *Theory and methods of scaling.* Wiley, 1958.

24. Whiting, J. W. M., & Child, I. L. *Child training and personality: a cross-cultural study.* Yale Univer. Press, 1953.

CHILDREARING ATTITUDES OF MOTHERS IN GERMANY AND THE UNITED STATES[1]

Don W. Rapp*
University of Illinois

Information concerning the relationship between childrearing attitudes in Germany and the United States has been, until recently, obtained largely from nonempirical data. Lewin (8), as an expert observer, expressed the view that German children are expected to exhibit unquestioning obedience, while children in the U. S. are urged to be more independent. Klineberg (6), however, has remained cautious in committing himself to statements regarding the German family's insistence on obedience. He used the phrase "allegedly characteristic" in reference to the reported widespread emphasis on obedience in the German culture. Supporting Lewin's views with direct cross-national investigations, Anderson (1), McGranahan (9), and Schaffner (11) used German and U. S. youth and young adults in detecting national differences with respect to a more highly developed authority structure in the German culture.

Cohn and Carash (2), using the California F Test, found a German sample of working men more "authoritarian" than a similar U. S. sample. The German sample had a higher authoritarian score on the F Test than any national sample hitherto reported. Kates and Diab (5) obtained a positive, although nonsignificant correlation, between the F Test and childrearing attitudes; i.e., the more authoritarian the more controlling the attitudes concerning children.

The present study sought to investigate directly whether the childrearing attitudes of German and U. S. mothers were empirically distinguishable.

Method

Subjects

German subjects were mothers from the Länder of Baden in southwest Germany, who responded to a German translation of the University of Southern California Parent Attitude Survey (12), PAS, and an attached personal information sheet. The sample was obtained through a Protestant

---

* Division of Child Development, University of Illinois, Urbana.

[1] A condensation of a dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Philosophy, Florida State Univer., August, 1960. The author wishes to express his gratitude to his graduate committee: Drs. Ruth J. Dales, Ruth Connor, Sarah Lou Hammond, Ralph L. Witherspoon, and James Walters.

Child Develpm., 1961, 32, 669-678.

kindergarten organization.[2] The investigator met with the 30 kindergarten teachers who served as liaison persons for the study. Distribution of 200 PASs was begun in May 1958, and in the same year 183 or 91.5 per cent were returned to the investigator in the U. S.

The PASs in the U. S. were distributed in 1960 to 28 Florida preschools in the same general manner as in the German sample. Of the 537 PASs distributed, 283 or 52.7 per cent were returned. The difference in the percentage of survey return between the German and U. S. samples was thought in part to be a function of the unifying organization to which the entire group of German teachers belonged. The U. S. teachers involved in this study lacked any common organizational ties.

The "kindergarten" in Germany includes a wider age range than the U. S. kindergarten; therefore the U. S. sample was drawn from mothers of kindergarten and nursery school children.

All subjects in both final samples were Caucasian, Protestant, and from intact homes. They reported parents, husbands, and themselves to be citizens of their respective countries, had at least one child in preschool, and completed PASs omitting no more than three items. Weights for omitted items were interpolated from the mean of the subscale to which they belonged.

The Index of Social Status (10), ISS, used in this study was originally based on U. S. research. Lefkowitz (7), testing its appropriateness for use with a European sample, found a correlation of $r = .82$ ($p < .01$) between expert local (Vienna, Austria) social class judgment and ISS categorical placement. The ISS determines the social status of a family by rating the occupation, source of income, and educational attainment of the "status parent." In this study information concerning the father was used. The ISS divides the social class continuum into five categories, two lower and two middle levels and one upper level. Subjects in the present study who fell into the ISS upper ($N = 4$) and upper-middle ($N = 41$) class categories were combined and designated as the "upper" class subsample with an N of 45. The subjects who fell into the ISS lower middle class category were designated as the "middle" class subsample of 41 subjects, and the subjects who fell into the upper-lower ($N = 30$) and lower-lower ($N = 8$) ISS class categories were combined into a "lower" class subsample of 38 subjects. The terms upper, middle, and lower as used in this study are relative terms and should be thought of in reference to their make-up. Social class breakdown, in ISS categories, was identical in both national samples. The upper, middle, and lower class subsamples, when considered separately, are designated as class samples; combined they are called complete samples.

The German and U. S. samples were equated by individually matching the U. S. surveys with the German surveys. Of the 124 German-U.S. pairs,

---

[2] Sincere appreciation to Pfarrer Herrenbrodt, head of the Gesamtverband der Inneren Mission, Karlsruhe, Germany, to Fräulein Gegenheimer, Mission supervisor of Baden, and to Fräulein Maria Fath, administrative secretary of the Mission, whose combined efforts made possible the collection of the German data.

86 were matched exactly for social class, age of mother in five-year intervals, and number of children in the family. The remaining 38 pairs were matched for social class, and as closely as possible on the other two categories, neither of which was found significantly different. The mean age of the German sample was 32.96, SD 5.58; for the U. S. sample, 32.80, SD 5.78. The mean number of children per mother for the German sample was 2.3, SD 1.23; for the U. S. sample, 2.5, SD 1.33.

The education of the mothers was not controlled in the equating of the two samples. In both German and U. S. samples 48 mothers reported two or more years of college; 59 U. S. and 19 German mothers reported themselves to be at least high school graduates or the equivalent; and 17 U. S. and 43 German mothers completed 11 years of school or less. Fourteen of the German mothers failed to report their educational attainment. The χ^2 of 31.1 summarizes the educational differences between the two samples.

Instrument

Shoben's Parent Attitude Survey was utilized in German and English forms. In Shoben's original research (12), the 85-item survey discriminated between mothers of problem children and mothers of nonproblem children. A low score has been interpreted as indicating a more optimum attitude concerning child guidance; a high score indicates the reverse. Three main subscales are present in the PAS: Dominant, Possessive, and Ignoring. The Dogmatism score (3) was derived from the actual number of extreme responses, i.e., strongly agree and strongly disagree.

The PAS was translated by four bilingual Germans.[3] The investigator was present with three of the four at the time of translation. Questions were answered concerning the particular American colloquialisms which frequent the PAS. These four translations were then used by a fifth translator[4] to bring together the final German form.

RESULTS

The complete samples were cross-nationally compared on the five scores obtained: PAS total score, Dominant, Possessive, and Ignoring subscale scores, and the Dogmatism score. All five tests were found significant beyond the .001 level with the U. S. sample demonstrating the smaller scores or the less controlling attitudes. Figure 1 presents graphically the wide differences found between the total score distributions of the two complete samples. The median test (13) χ^2 value was 80, significant well beyond the .001 level. This large difference is reflected in the statement: only 4 per cent of the U. S. mothers had total scores above the median of the complete

---

[3] Dr. G. Hassfurther, Dr. R. Defieber, Dr. H. Reinstrom, and Dr. M. Suessman.

[4] Dr. Marlies Kronegger.

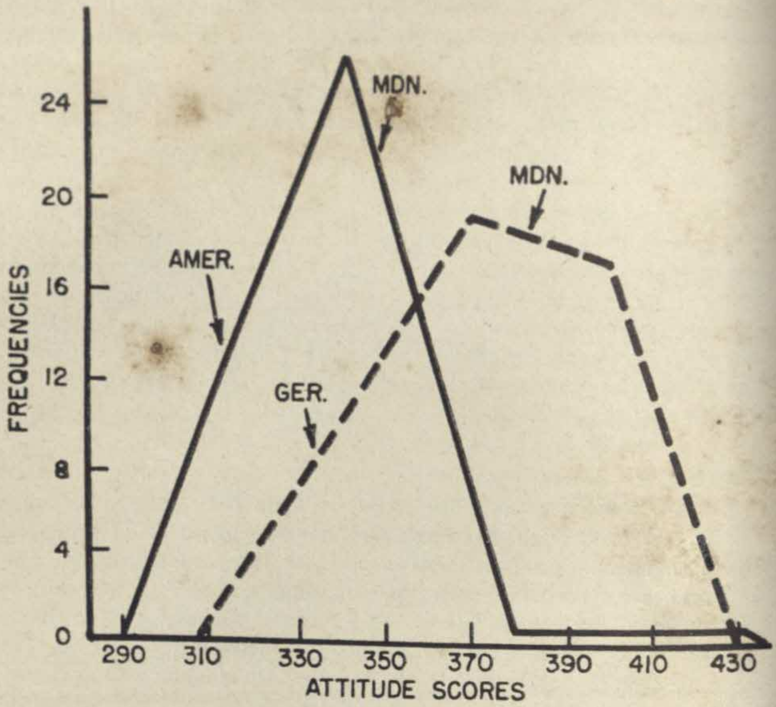

FIGURE 1—Distributions of total PAS scores from the complete German and United States samples. The median test χ^2 value was 80 ($p < .001$).

German sample. Other values on the complete sample cross-national tests were: Dominant subscale ($\chi^2 = 62.02$), Possessive ($\chi^2 = 76.94$), Ignoring ($\chi^2 = 12.75$), and Dogmatism ($\chi^2 = 19.76$). Considering the great disparity between the German and U. S. samples on all five scores tested, it can be said that the two samples were most probably drawn from attitudinal populations with different medians.

German and U. S. comparisons of like social classes—i.e., German upper-class to U. S. upper-class, etc.—were made using the Mann-Whitney U test (13). Corrections were made for ties. Figure 2 presents these comparisons in horizontal relationship on all five scores tested. All of the three U. S. social classes were found to possess smaller PAS median scores, indicative of less controlling childrearing attitudes, than their German social class counterparts. Thirteen or 87 per cent of the 15 comparisons reached significance, one at $p < .05$ and 12 at $p < .01$. The two tests not reaching significance were the cross-national upper class comparison in the Ignoring subscale and the middle class Dogmatism comparison. They were, however, in the expected direction ($p < .10$, $p < .17$, respectively). The findings

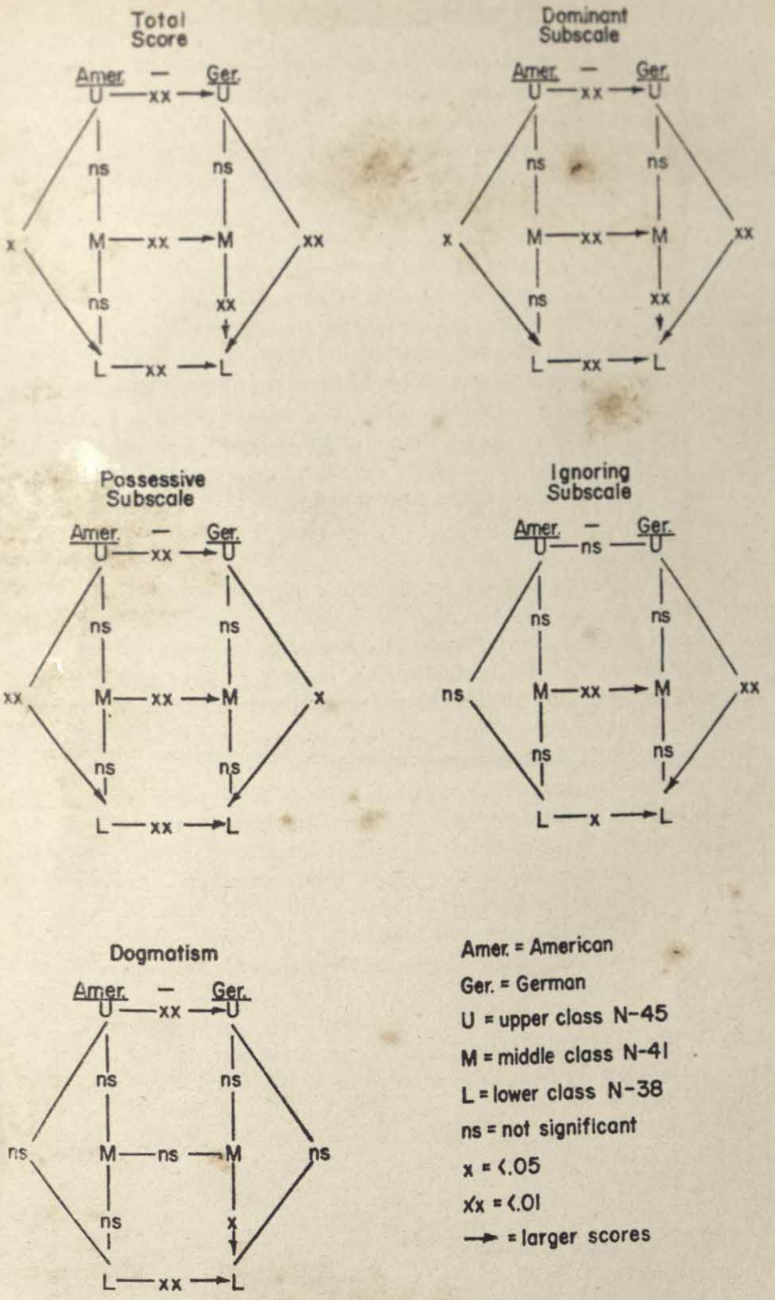

FIGURE 2—Inter- and intranational social class relationships of German and United States childrearing attitudes.

leave only a minimum of doubt that comparable German and U. S. class samples were independent samples drawn from separate populations.

Intranational class comparisons were computed between upper and middle, middle and lower, and upper and lower class samples on all five scores obtained. Figure 2 presents the intranational comparisons in vertical relationship. Twenty-seven of the 30 class comparisons indicated the following pattern: upper class scores smaller than middle and lower class scores, and middle class scores smaller than lower class scores. Ten or 33 per cent of the 30 intranational tests were found significant, four at $p < .05$ and six at $p < .01$. Significance was generally reached between upper and lower class samples in both national samples and between the German middle and lower class samples. Upper and middle class samples in both national samples responded quite similarly. Thus, it is evident that the U. S. social classes exhibited less variability than did the German social classes. The reason is primarily due to the significant differences reached between the German lower and middle classes as opposed to no U. S. differences in this relationship.

Discussion and Conclusions

A major limitation in all research crossing a language barrier is the insecurity of identical bilingual understanding of the material presented. A second limitation should also be noted. The German situation made it impossible to pretest the German translation. The manner in which the survey was translated, however, was designed to overcome these limitations to the greatest degree possible.

The PAS in the original study discriminated between mothers of problem children and mothers of nonproblem children. German scores approximated the scores obtained by mothers of problem children, and the U. S. scores approximated the scores obtained by mothers of nonproblem children. This should not be interpreted as indicating that German children have a greater chance of becoming problem children than U. S. children. In this respect the uniqueness of culture does not allow judgmental interpretation.

The German mothers reported significantly less education than the U. S. mothers. The matching inaccuracy stems from the fact that the ISS determines social class by using criteria based on the husband. The relationship indicated here is an interesting one. The educational level of the U. S. mothers in this sample was similar to that of their husbands, while the educational level of the German mothers was much less than their husbands. Relevant to the present findings, it has been found (12, 15, 17) that the more education a person has the less controlling their childrearing attitudes. It would then seem possible that part of the large score difference found in the present study could possibly be attributed to the difference in educational level of the two national samples. Analysis of this possibility was conducted as follows: The total scores of the U. S. lower class (least education of the U. S. sample) were compared with total scores of the German upper class (most education in the German sample). Differences were still found,[5] indicating the U. S. lower class had significantly smaller scores or less controlling attitudes than the German upper class.

The three U. S. social classes exhibited less variability than did the German social classes (Figure 2). These data are interpreted as a reflection of cultural influences. Specifically, greater social class uniformity in the U. S. sample is seen, at least in part, as the result of the American governmental system advocating equality and the public educational system which promotes contact between social classes. Greater German social class disunity is seen partly as a result of the very recent dictatorial militaristic form of government and a more authoritarian class determining educational system. This finding is in support of McGranahan's (9) statements indicating that German-U.S. differences in attitudes are not peculiar to political attitudes, but spread throughout a wide variety of attitudinal areas into even the simplest types of social relationships.

Comparison of PAS Scores

Table 1 has assembled the means and SDs of several studies which have used the complete form of the PAS.[6] The value of placing present study results in the context of previous investigations is beneficial not only to the present research but allows multiple comparison. The German total and Dominant mean scores are considerably higher than all the scores presented except for Shoben's so-called problem group. As shown in this study, the U. S. sample demonstrated significantly smaller scores than the German sample. However, the U. S. sample also has total and Dominant means that are considerably higher than all but those for Shoben's problem group.

Specifically, all three of the PAS means presented by Walters *et al.* (16) were obtained from samples of mothers who generally belonged to the upper middle class. It was thought possible that this class restriction in the "three state sample" (16) may have been one of the reasons why the present study U. S. sample, with its full range of social class, tended to exhibit higher scores. However, when the mean of the upper class U. S. sample (344.2, SD 15.9, N 45) was compared with the mean of the Florida, Alabama, and Connecticut means (328.1, approximated SD 26,[7] N 90), a significant difference was found well beyond the .01 level. Direction and

---

[5] Mann-Whitney U test $z = 3.67$; $p < .001$. Two-tailed tests of significance were used throughout the analysis.

[6] Only 84 of the 85 PAS items were used in this study. Item 1 of the Ignoring subscale was omitted from the scoring of both samples. A typographical error on the German form rendered it dissimilar in meaning to the English version. In order to make these data of the present study comparable to the other PAS data presented, four score points (an approximation of the mean weight responses on the omitted item) have been added to the total and Ignoring scores in both samples.

[7] The approximated SD of 26 is presumed a reasonable if not a high estimate of the true SD considering the homogeneous social class nature of the "three state sample" and also considering the range and magnitude of the other SDs presented in Table 1.

TABLE I

MEANS AND STANDARD DEVIATIONS OF PAS SCORES OBTAINED IN COMPARABLE STUDIES

	Date of Data Collection	TOTAL		DOMINANT		POSSESSIVE		IGNORING	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Present Study									
American mothers	1960	†352.1	26.0	169.3	13.1	82.8	8.4	† 56.5	5.2
German mothers	1958	385.3	24.9	186.6	13.9	94.7	7.7	† 59.5	5.1
Shoben									
Nonproblem mothers (Gp I)	1949*	324.7	21.7	154.7	11.6	91.1	6.4	42.8	4.2
Problem mothers	1949*	397.7	29.4	187.8	16.7	111.7	10.3	56.1	5.9
Clinical psychologists	1949*	286.4	6.7	146.1	2.2	79.4	4.6	39.9	2.7
Freeman									
Nonschiz. mothers	1955*	. . . .	. . .	158.5	18.1	85.9	10.7	56.8	5.4
Schiz. mothers	1955*	. . . .	. . .	161.0	16.7	89.6	11.3	58.6	6.7
Walters et al.									
Florida mothers	1957	322.2	. . .	154.6	. . .	78.0	. . .	54.4	. .
Alabama mothers	1955	333.7	. . .	157.3	. . .	77.3	. . .	55.6	. .
Connecticut mothers	1957	328.6	. . .	156.7	. . .	77.2	. . .	54.0	. .
Walters & Fisher									
September, 1953	1953	338.3	. . .	158.0	. . .	82.8	. . .	54.9	. .
January, 1954	1954	326.9	. . .	152.0	. . .	78.9	. . .	59.6	. .
November, 1955	1955	311.4	. . .	146.0	. . .	76.4	. . .	49.7	. .
Walters & Bridges									
Men: Average of seven college samples	1955	337.7	. . .	159.6	. . .	82.9	. . .	52.2	. .
Women: Average of two college samples	1957*	338.6	. . .	160.0	. . .	82.6	. . .	54.3	. .

* Data collected prior to date indicated.

† Four score points have been added in order to render the means comparable to the other means presented (*see* footnote 6).

magnitude of this comparison tends to lower the probability of social class completely explaining the mean score difference.

U. S. data of the present study were collected in 1960 while data from other American studies were collected in 1957 or before. The present study U. S. total and Dominant subscale scores appear higher than any of the other mean scores here presented, with the exception of that for Shoben's problem mother sample. The question arises, "Within the last few years are the mothers in the United States becoming more dominant and controlling over their children?" If this is evidence of a general shift to a more controlling childrearing attitude, it is one that coincides with the coming of the sputnik era with its re-emphasis on the three R's and accompanying public re-evaluation of educational techniques and curricula.

SUMMARY

The purpose of this investigation was to determine the measurability and relationship between German and U. S. childrearing attitudes. Parent Attitude Surveys from 124 Florida mothers were individually matched with translated surveys completed by 124 German mothers from Baden in southwest Germany. Matching was near perfect on three categories, social class, age, and number of children in the family.

Significant differences were found between the German and U. S. complete samples on all five scores obtained. The U. S. mothers demonstrated scores indicating less controlling attitudes than the German mothers. These differences were maintained in 13 of the 15 cross-national comparable class comparisons. The analysis within social class showed similar relationships in both samples. Upper and middle class scores were very much the same within each national sample; however, the lower classes demonstrated the most controlling and authoritarian attitudes within their respective cultures. Social classes in the U. S. were found attitudinally less variable than those in Germany with respect to childrearing attitudes.

A comparison of available Parent Attitude Survey scores from other studies was presented.

REFERENCES

1. ANDERSON, H. H., ANDERSON, G. L., COHEN, I. H., & NUTT, F. D. Image of the teacher by adolescent children in four countries: Germany, England, Mexico, United States. *J. soc. Psychol.*, 1959, 50, 47-55.
2. COHEN, T. S., & CARSCH, H. Administration of the F scale to a sample of Germans. *J. abnorm. soc. Psychol.*, 1954, 49, 471.
3. FREEMAN, R. W., & GRAYSON, H. M. Maternal attitudes in schizophrenia. *J. abnorm. soc. Psychol.*, 1955, 50, 45-52.

4. Jahoda, M., Deutsch, M., & Cook, S. W. *Research methods in social relations*. Dryden Press, 1958.

5. Kates, S. L., & Diab, L. M. Authoritarian ideology and attitudes on parent-child relationships. *J. abnorm. soc. Psychol.*, 1955, 51, 13-16.

6. Klineberg, O. Cultural factors in personality adjustment of children. *Amer. J. Orthopsychiat.*, 1953, 23, 465-471.

7. Lefkowitz, M. Authoritarianism in adolescents: a cross-national study. Unpublished doctoral dissertation, Univer. of Texas, 1956.

8. Lewin, K. *Resolving social conflicts*. Harper, 1948.

9. McGranahan, D. V. A comparison of social attitudes among American and German youth. *J. abnorm. soc. Psychol.*, 1946, 41, 245-257.

10. McGuire, C., & White, G. D. The measurement of social status. Research paper in human development, No. 3 (Rev.). Univer. of Texas, Austin, 1955.

11. Schaffner, B. *Fatherland: a study of authoritarianism in the German family*. Columbia Univer. Press, 1948.

12. Shoben, E. J., Jr. The assessment of parental attitudes in relation to child adjustment. *Genet. Psychol. Monogr.*, 1949, 39, 101-148.

13. Siegel, S. *Non-parametric statistics for the behavioral sciences*. McGraw-Hill, 1956.

14. Walters, J., & Bridges, B. Attitudes of single men toward child guidance. *J. home Econ.*, 1956, 48, 109-113.

15. Walters, J., & Fisher, C. Changes in the attitudes of young women toward child guidance over a two-year period. *J. educ. Res.*, 1958, 52, 115-118.

16. Walters, J., Strickland, M. P., & Christopherson, V. A. Paternal and maternal attitudes concerning the guidance of children. Unpublished manuscript, Dept. Home & Family Life, Florida State Univer., 1959.

17. Zuckerman, M., Barrett, B. H., & Bragiel, R. M. The parental attitudes of parents of child guidance cases. *Child Develpm.*, 1960, 31, 401-417.

NEW TECHNIQUES FOR RESEARCH IN CHILDREN'S TEMPTATION BEHAVIOR

ROBERT E. GRINDER*
University of Wisconsin

The observation that progress in a scientific field usually waits upon the discovery of a satisfactory dependent variable (12) seems particularly relevant in regard to temptation behavior. The term "temptation" is used here to refer to a type of conflict where a choice must be made between (a) conformity to some socially expected role behavior and nonattainment of a goal and (b) nonconformity and attainment of the goal.

For several years, on interview, case history material, and theoretical bases, personality theorists have viewed the kind of response(s) that an individual makes to temptation as one of the indices of the construct conscience or superego. For example, Sears *et al.* (11) suggest that capacity to resist temptation is one of the consequences of conscience development. Freud speaks of the superego as generating moral anxiety as it "observes, guides, and threatens the ego" (6, p. 89). Fenichel describes the "warning function" of conscience and distinguishes guilt feelings from "feelings which do not judge the past but the future" (4, p. 134). Redl and Wineman, in discussing "value danger signals," suggest that the superego uses the production of feelings of guilt as either "pre-situational or post-situational signals of value conflict to the ego" (10, p. 200). The contribution of these and similar studies to theoretical knowledge and empirical ways of assessing the conscience or superego construct is valuable; therefore, it is necessary that operational definitions of responses to temptation be developed in order to test them as dependent variables in empirical investigations of superego functions. Because research method in this area is by and large uncharted, this paper attempts to outline some of the general theoretical problems and to provide a methodology which has been demonstrated as a measure of children's temptation behavior (3, 7).

THEORETICAL FRAMEWORK

The common-sense notion of temptation is offered by the Oxford Dictionary (13) which presumes a person to be tempted when he is "enticed to do what is wrong by the prospect of some pleasure or advantage." For empirical research, however, it seems essential that the meaning of the word be specified in a strictly functional or behavioral sense (cf. Bijou's [2]

---

* School of Education, University of Wisconsin, Madison 6.

Child Develpm., 1961, 32, 679-688.

approach to an empirical behavior theory for children). A definition of temptation of this kind may be conceptualized as follows: a state of temptation exists for an individual when the addition or presence of a rewarding stimulus (incentive) increases the probability of a class of responses which is forbidden because of its incompatibility (conflict) with some socially expected behavior.[1] This definition is based upon the functional relationships between the elements of a temptation situation itself.[2] The incentive for temptation conceivably might be an intangible such as the anticipation of praise or the circumventing of anticipated reproof, or a tangible such as money, candy, a toy, or freedom from corporal punishment. The behavior upon which transgression is contingent might be evaluated from the point of view of virtues, e.g., honesty, truthfulness, punctuality, responsibility, etc., or from that of tabus, e.g., cheating, lying, procrastination, carelessness, etc. According to the definition, then, a tempted individual is in the sort of conflict that Miller (9) has named "double approach-avoidance"; both alternatives, either yield to temptation or resist temptation, appear as approach tendencies, but the necessity of relinquishing one of the alternatives in order to pursue the other may instigate a tendency to avoid the one which is approached.

Whatever the character of the conflict, the probability of transgression, and thus the degree of conflict, (a) will vary inversely with extent of coercion and/or the possibility of being detected and will vary directly with (b) the extent to which the expected behavior has been learned and (c) the incentive value of the tempting objective (cf. below). Further, (d) the objective of temptation may be either an acceptable goal which can be obtained only by means which are forbidden or a goal which is forbidden. Finally, (e) temptation conflict, in the sense that the person is pushed in two or more directions simultaneously (5), ceases when the desired goal has been achieved or when it has been successfully resisted.

From the foregoing presentation it may be concluded that temptation is a predecision-type conflict situation where an individual, free from coercion or danger of detection, must choose between a positive incentive and conformity to learned role behavior.

If subjects are to be compared with one another in the degree to which they either resist (conform) or yield (transgress) to temptation, it is necessary that the variables of the temptation conflict be measurable indices. The expected behavior pattern provides a single continuum upon which the variables can be ordered. If a subject conforms to the pattern, i.e., if he behaves as he would under nontemptation conditions, he can be scored as having resisted temptation. There is, in other words, zero discrepancy between his behavior and the expected behavior. On the other hand, if he transgresses, his responses of deviation can be taken as measures of yielding to temptation. There is, in this instance, a discrepancy between the subject's behavior and the expected behavior.

---

[1] Socially expected behavior is conceptualized to be a specific aspect of role behavior, e.g., a child assigned to be class monitor might transgress this role by throwing a spitball, deserting his post, or napping, etc. In an experiment socially expected behavior may be defined by consensus of qualified judges.

[2] It is recognized nonetheless, that in seeking developmental antecedents of temptation behavior, as for example in aspects of socialization (1, 3, 7), it might be useful to postulate personality constructs (e.g., conscience or superego) or to infer such motivational properties as drives, desires, wishes, etc. Further, on this level of analysis, it might be appropriate to introduce such intervening constructs as identification, introjection, habit, etc.

These indices also are not independent of each other. The conflict is a function of the attraction of the tempting goal *and* the extent to which the expected behavior has been learned. Consider the problem, important to studies of self-control, of comparing subjects in terms of resistance to temptation: some subjects may appear to resist temptation whereas actually the incentive is not tempting, and others may apparently transgress, not because they have not learned the expected behavior as well as those who have conformed, but because the incentive has relatively more appeal to them (7). Obviously, only when the tempting object has equal incentive value to all subjects can the extent to which they conformed be meaningfully compared.

Controlling the incentive value of the tempting goal is difficult. The tempting objective, although on occasion intangible, is usually some tangible, manipulable goal, the incentive value of which must be inferred from goal-directed behavior in nontemptation, nonconflict situations. Also, the relative desirability or reward properties of goal objects, e.g., candy, toys, or money, etc., must be demonstrated for each individual. The great difficulty of these tasks is illustrated by the fact that in the now classical Hartshorne and May (8) Character Education Inquiry efforts to control temptation were abandoned. In the methodology presented here, a simple procedure for controlling the incentive value of the tempting goal proved itself to be useful. The same incentive stimuli were offered to all subjects in an experiment, and an attempt was made to make the incentive attractive so that every subject's motivation would be high.

In addition to settling upon suitable indices of the dependent variables and assessing temptation to transgress, in devising an experimental task a number of technical considerations are necessary. First, so that activity in the situation will be appealing, interesting, provocative, meaningful, etc., experimental circumstances should be appropriate to the subjects' sophistication. That is, the activity should be sufficiently complex as to arouse no suspicion or fear of external coercion and/or detection, while at the same time transgression is relatively easy. Second, achievement or progress toward the goal should be constant in order to assure that the tempting goal is equally attainable by all subjects. Otherwise, some subjects, skilled in the task, may not be tempted to transgress because the goal can be attained readily. Third, since temptation to transgress may be contingent upon knowledge of the goal's presumed attainability, subjects should have opportunity to keep a cumulative record of their progress toward the goal.

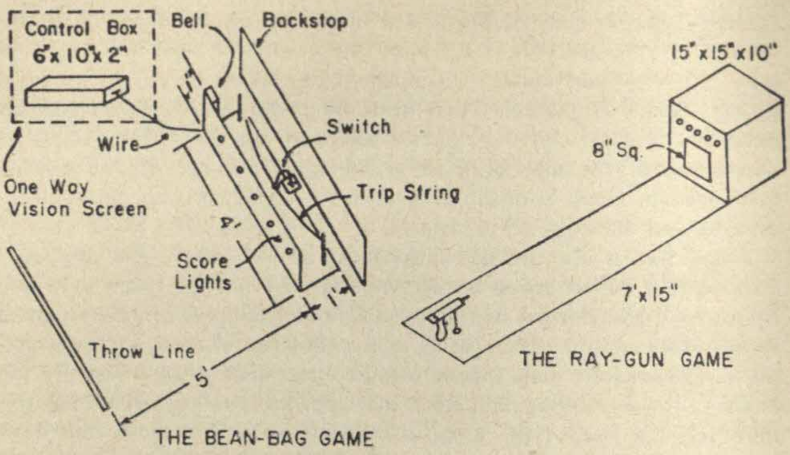

FIGURE 1—The apparatus for measurement of temptation behavior.

Fourth, an objective recording of the subjects' behavior should be made. Fifth, the apparatus should be of a sort that is easy to transport and operate, because of practical difficulties in bringing subjects to a laboratory and because fear of detection is minimized in a familiar setting.

With these considerations in mind, two temptation situations have been devised, a beanbag game for children 4 to 5 years old and a ray-gun game for those from 6 to 15 years old (Figure 1).[3] A game situation was chosen for both age groups for several reasons. First, it seemed likely that each child would have learned in varying degrees that in a game his behavior should be to "play fair" or to "play according to the rules." Second, the expected behavior can be specified in rules easily understood by children. Finally, it is likely that to some extent this aspect of socialization would have been introduced relatively early in the child's training. The measurement procedure for both age groups generally is the same. Subjects, asked to play individually, are tempted to transgress its rules in order to win an appealing incentive.

METHODOLOGY

The Beanbag Game

Procedure. A subject is individually instructed by an experimenter to stand behind, and not to cross, a throw-line drawn parallel 5 ft. from a 1-

---

[3] The methodology presented here and the essential ideas for the construction of the beanbag game grew out of discussions at a graduate seminar at the Laboratory of Human Development at Harvard. While the staff of the Laboratory and other members of the seminar should not be held responsible for this particular formulation, the basic ideas are a result of our joint effort. The writer wishes to acknowledge in particular the guidance received from John W. M. Whiting, Wesley Allinsmith, Eleanor E. Maccoby, and Roger V. Burton.

by 4-ft. panel upon which five 1-in. pilot (score) lights are evenly spaced across the front. The subject is given five beanbags which he is to toss one at a time over the panel. If a bag "scores," it turns on a light. None of the bags is to be retrieved. The subject is told that if he can turn on three lights he will win an exciting toy; he is then shown a bowl full of 25 cent toys from which he is asked to choose the one he would like most to have if he should win.

To insure that the subject understands the rules, he is asked to play a practice game. As the next game, the one for the toy, is about to begin, the experimenter announces that he will have to attend to a telephone call. The subject is then asked (because the experimenter will have to be away for a few minutes) to play the game by himself. The scores each subject earns by tossing the bags are controlled by an observer behind a one-way vision screen. If the subject follows the rules, only one light will register. However, if he breaks the rules, a light will register on each occasion that an unauthorized bag is tossed. The subject is judged to have resisted temptation if he conforms to the rules; he is judged to have yielded to temptation if he transgresses. The latter may be measured in terms of how many bags are tossed before the subject transgresses or in terms of the number of extra tosses or the number of times the throw line is crossed, etc.

After sufficient time has elapsed for the child to play the game, the experimenter returns to the game room. At this point, at least two alternatives are possible. He may express regret that the child who conforms did not win a toy, and he may reward the child who has transgressed with the toy of his choice. On the other hand, as Burton *et al.* (3) chose to do, he might suggest that the game be played again for the toy. Now the observer turns on four lights as the bags are tossed. Every child departs with a toy, and an opportunity to relieve possible guilt is offered to those who have transgressed.

Technical description. Figure 2 shows the schematic circuit of the beanbag game. The dashed line separates the control box from the game. A multiconductor cable, which runs along the floor to behind the one-way vision screen, connects the two units. Notice that the control box houses a circuit on-off switch (S_1), a six position rotary-shorting switch (S_2) for the score lights, a push-button switch (S_3) for ringing a chime every time a score light registers, a score circuit switch (S_4) parallel to a similar switch (S_5) on the front panel, and a score indicator lamp. The game, built of plywood, is approximately 4 ft. long with the panel and its base about 12 in. in height and width, respectively. The five score lights are evenly spaced along the bottom of the front panel. A switch (S_5), mounted in the upper left-hand corner of the front panel, enables the experimenter to turn off the score lights after an experimental sequence. When S_5 opens the circuit, the observer at the control box can return the rotary switch "home," and then close the circuit again with S_4. A string, which passes through a loop on the end of a stem of a microswitch (S_6), serves as a target for the

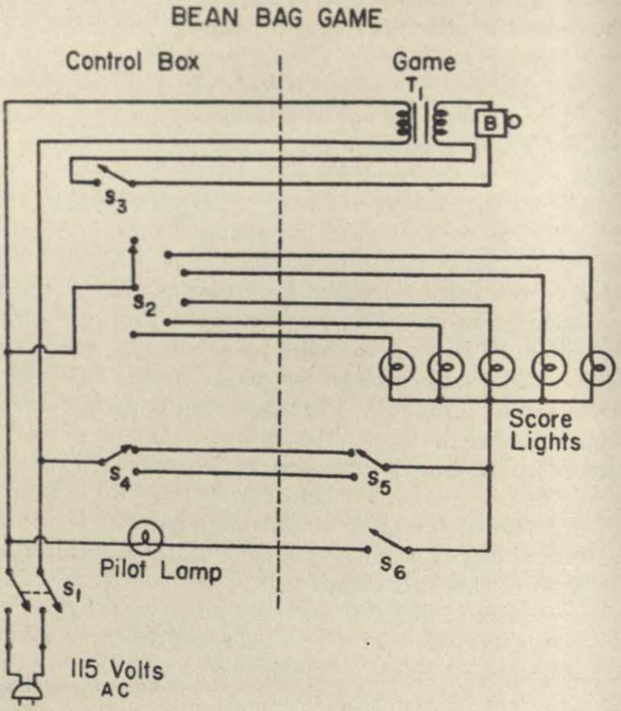

FIGURE 2—Schematic circuit of the beanbag game. S_1 = D.p.s.t. toggle. S_2 = 6-pole 1-position rotary shorting switch. S_3 = S.p.s.t. push button. S_4, S_5 = S.p.d.t. toggle. S_6 = S.p. momentary position microswitch. T_1 = Bell transformer, 115 v., 6.3 v. B = Doorbell, 6.3 v. Bulbs = 6 w., candelabra base, 115 v.

beanbag tosses. Whenever the switch is momentarily closed by a bag hitting the string, a score indicator lamp on the control box flashes. The string, mounted behind the front panel and out of sight of the subject, runs the length of the game; S_6 is mounted in the center of the base. Several modifications of the apparatus are obviously possible. Burton *et al.* (3), for example, substituted SPST mercury switches wired in series for the control box rotary switch in order to eliminate its contact noises; also, they built a portable one-way vision screen which made it possible to use the game in any available school room.

The Ray-Gun Game

Procedure. The game consists of a "ray-gun," a 7 ft. by 15 in. plank (hinged together in three sections) upon which the gun is swivel mounted, and a 15 by 15 by 10 in. plywood box. On the front of the box are two revolving illuminated rockets set behind a piece of translucent glass, 8 in. square and five 1-in. pilot lights evenly spaced across the front of it above the target glass. Scores from 0 to 5 are registered on the pilot lights with each pull of the trigger of the ray-gun. Obtained scores are prearranged (cf. technical description below).

The experimenter demonstrates, for about 10 subjects at a time, the game and the rules for playing it. Subjects are told that if they "hit" the center of one of the rockets they will score a bull's-eye and all five of the score lights will shine; if they hit the head or the tail of the rocket they might get a score from 0 to 5 depending on how close they are to the center. Subjects are asked to take 20 shots and to record their scores on a three-column, 20-row score sheet with a ball point pen. The first column, labeled "shot number," lists the row number; the second column, labeled "score obtained," is blank in order that a subject can record whatever score he might choose for each shot; the third column, labeled "total so far," is for a subject to record cumulatively his scores.

Subjects are told that if they obtain from 35 to 39 points they will win a Marksman badge, 40 to 44 points a Sharpshooter badge, and 45 points or more an Expert badge. As has been noted, scores are prearranged, and the experimental sequence totals 32 points when *only* obtained scores are recorded. The badges, 1¾ in. in diameter, were prepared by a commercial badge firm which engraved in black a drawing of the ray-gun upon a background colored differently for each of the three types of badge.[4]

After explaining the rules, the experimenter "aims" carefully, takes a few practice shots, and records his scores. Subjects are then returned to their classroom. Each is taken out of the classroom individually, led to the experimental room, and at the door is told that the experimenter is busy working on some "things" down the hall. The subject is asked to play the game alone and to bring his score sheet to the experimenter when he finishes. The door of the experimental room is then closed, and the experimenter audibly walks off.

The subject is judged to have resisted temptation if it appears he has conformed to the rules, i.e., if his score sheet shows a total score of 32 points. He is judged to have yielded to temptation if his score sheet indicates that he has altered the prearranged sequence. Transgression may be measured either by one of several highly correlated variables, e.g., the number of times a low score is skipped, the number of times a higher score than the obtained is recorded, the total score, etc., or by a weighted combination of scores, e.g., a discriminate function score.

When the subject brings his score sheet to the experimenter, he is thanked for playing the game. If his score sheet indicates conformity, regret is expressed that he has not won a badge; if it shows that the subject has won a badge, it is pinned on his shirt. No attempt is made to reward those who had conformed or to relieve any possible guilt of those who had transgressed. When playing the game twice (following the procedure of the beanbag game) was tried in pretesting, it was found that the news that a badge could be won on the second try was communicated to subjects who had not yet played the game. This knowledge effectively eliminated temptation the first time they played the game.

---

[4] The influence of peer-group status upon temptation to transgress must be investigated before this kind of incentive is employed extensively.

Technical description. Figure 3 shows its schematic circuit. Scores are automatically prearranged (thus eliminating the necessity of an observer) with two 20-pole step-in relays (Ry_1-Ry_2) wired in series with Ry_8. An impulse circuit transmits current, for a few millisec., to the combination relay whenever the ray-gun trigger is pulled. The impulse circuit prevents rotation of the relay when the trigger is held down while at the same time moving the relay to a new position when it receives an impulse. The output of the combination step-in relay is fed into a program board, constructed with two 20-terminal molded bakelite strips, which controls the score light lock-in relays (Ry_4-Ry_7). The lock-in relays can be wired to the step-in relays so that at each pole any number of score lights from 0 to 5 will register. The number of lights which are lit signify the score for the immediately preceding shot. A score continues to register until the next shot is taken. The total number of shots which a subject takes is recorded with an impulse counter (C).

The ray-gun is fashioned by inserting a soldering gun trigger switch (S_1) in the handle of a pistol-type acetylene torch. The barrel of the pistol is encased in the center of a circular ¾-in. long, ⅞-in. diameter piece of relatively hard rubber which is itself inserted in a ¾-in., ⅞-in. inside diameter piece of water pipe. A 5-in. rod tapped into the water pipe is screwed into the base of the plank and provides a swivel for the ray-gun. This prevents it from being pointed upward or downward, i.e., way off target, while at the same time scores are being registered. A cord running under the plank connects the gun to the game.

The target assembly is built so that two small rockets appear a total of 64 times a minute behind the 8 in. square translucent glass. A 7¾-in. circular template, with two holes, 1 in. in diameter, drilled opposite to each other, 6 in. apart, is centered upon the shaft of a midget 4 rpm motor (M). The motor is situated so that the template rotates behind the translucent glass. A piece of black drawing paper covers the inside surface of the glass. Eight 1-in. holes, evenly spaced in a circle to correspond with the two holes in the rotating template, are cut in the drawing paper. The holes allow light to pass through whenever the holes are matched. A small facsimile of a space-rocket is glued in the center of each template hole. Red celluloid covers the drawing paper so as to provide a glowing red background for the rockets. A 6-w. bulb mounted behind each of the holes serves as a projector lamp. In order that the game will be all the more exciting, the lamps are wired to blink as the ray-gun trigger is pulled.

A push button switch (S_2) situated at the back of the game box allows the experimenter to return the combination step-in relay "home," i.e., to the beginning of the experimental sequence. Note that a cycle of 40 scores appears before the sequence repeats itself; thus, it is extremely difficult for a subject who takes several extra shots to suspect a pattern. Transgression is presumed to be facilitated by programming zero scores at the beginning and at the end of the 20-shot pattern and by following high scores with zero scores in the middle of the sequence.

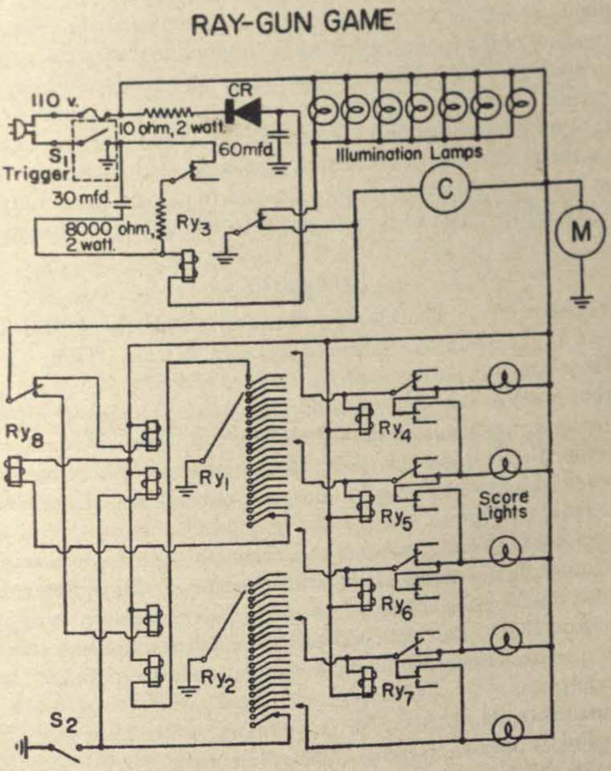

FIGURE 3—Schematic circuit of the ray-gun game. Ry_1-Ry_2 = Midget electrical reset stepper, 115 v., 20-pole. (Guardian MER 115). RY_3-Ry_7 = D.p.d.t. electrical relay, 115 v. (Potter and Brumfield GA11A). Ry_8 = S.p.d.t. electrical relay, 115 v. (one pole of P&B GA11A). S_1 = The ray-gun trigger (*see* text). S_2 = S.p.s.t. push button. C = Electric 3-digit reset counter, 115 v. (Veeder Root U-15073). M = Midget 4-rpm motor, 115 v. (Telechron). CR = Contact rectifier, 100 ma. (Federal selenium rectifier 1004A). Bulbs = 6 w., candelabra base, 115 v.

SUMMARY

This paper has (a) set forth some of the theoretical problems in the measurement of children's temptation behavior and (b) presented a methodology which has been demonstrated to be useful in its measurement. The problems and procedures are discussed in the belief that they may be used, in either this or a modified form, for further empirical research on conflict and conscience. Temptation is conceptualized as existing for an individual when the addition or presence of a rewarding stimulus (incentive) increases the probability of a class of responses which is forbidden because of its incompatibility (conflict) with some socially expected behavior. Settling upon suitable indices of the conflict, assessing temptation to transgress, minimizing fear of detection, and holding constant goal attainability appear as some of the major problems in temptation behavior measurement. A beanbag and a ray-gun game situation have been devised, and research (3, 7) indicates that the procedures are successful. The basic methodology is essentially the same in both of the games. Subjects, asked to play the game individually, are tempted to transgress its rules in order to win an appealing incentive.

REFERENCES

1. ALLINSMITH, W. The learning of moral standards. In D. R. Miller & G. E. Swanson, *Inner conflict and defense.* Holt, 1960. Pp. 141-176.
2. BIJOU, S. W. Learning in children. *Monogr. Soc. Res. Child Develpm.,* 1959, 24, No. 5 (Whole No. 74), 25-36.
3. BURTON, R. V., MACCOBY, E. E., & ALLINSMITH, W. Antecedents of resistance to temptation in four-year-old children. *Child Develpm.,* 1961, 32, 689-710.
4. FENICHEL, O. *The psychoanalytic theory of neurosis.* Norton, 1945.
5. FESTINGER, L. *A theory of cognitive dissonance.* Row, Peterson, 1957.
6. FREUD, S. *New introductory lectures on psychoanalysis.* Norton, 1933.
7. GRINDER, R. E. Parental childrearing practices, conscience, and resistance to temptation of sixth grade children. *Child Develpm.,* in press.
8. HARTSHORNE, H., & MAY, M. A. *Studies in the nature of character. Vol. I: Studies in deceit.* Macmillan, 1928.
9. MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders, Vol. I.* Ronald Press, 1944.
10. REDL, F., & WINEMAN, D. *Children who hate.* Free Press, 1951.
11. SEARS, R. R., MACCOBY, E. E., & LEVIN, H. L. *Patterns of child rearing.* Row, Peterson, 1957.
12. SKINNER, B. F. *Cumulative record.* Appleton-Century-Crofts, 1959.
13. *Oxford universal dictionary on historical principles.* (3rd Ed., Rev. with addenda) Oxford Univer. Press, 1955.

ANTECEDENTS OF RESISTANCE TO TEMPTATION IN FOUR-YEAR-OLD CHILDREN[1,2]

ROGER V. BURTON,* ELEANOR E. MACCOBY,[3]
and WESLEY ALLINSMITH[4]
Harvard University

This study is in the tradition of many recent investigations which explore factors in the development of personality by relating childrearing practices to later behavior in the child. As part of the larger question of how one internalizes standards and rules of behavior, we ask: how does a person learn to restrain himself from doing something he very much wants to do but which his socializers have tried to teach him not to do? In this investigation, the type of resistance to temptation studied was conformity to the rules of a game. The focus will be on the relations found between childrearing practices as reported by the mother and the child's behavior during an experimental test of resistance to temptation.

Hypotheses derived from identification theories as formulated by Allinsmith (1), Maccoby (18), Sears, Maccoby, and Levin (22), and Whiting

---

* Laboratory of Socio-environmental Studies, National Institute of Mental Health, Bethesda 14, Maryland.

[1] We wish to express our appreciation to John W. M. Whiting whose generous help and creative suggestions have contributed to every stage of this study. Thanks are also due Robert E. Grinder, who contributed to the development of both the bean bag game and the total experimental design, and to the staff of the Laboratory of Human Development, Harvard University, who aided in gathering the data. We also thank Drs. John D. Campbell, John A. Clausen, Thomas L. Gillette, and Marian Radke Yarrow for constructive comments on early drafts of this paper.

[2] The main support for this study was provided by a grant from the National Institute of Mental Health (M-1096).

[3] Stanford University.

[4] University of Cincinnati.

Child Develpm., 1961, 32, 689-710.

(27, 29) mainly determined the selection of childrearing practices to measure. These formulations, taking much from psychoanalytic writings, propose that the child adheres to parental standards as a consequence of an identification process whereby the child patterns himself upon a parent. Two motives are generally posited for this process: the desire to reproduce the gratification provided by a loved object and the desire to control, neutralize, or protect oneself from a fearful, threatening, or powerful object. Whiting (29) suggests that these two kinds of identification are really based on the single motive of envy of the controller of resources.

Another body of psychological theory based on the effects of direct reward and punishment provides parallel and sometimes different hypotheses concerning origins of resistance to temptation. According to this general behavior theory a concomitant of punishment is fear which becomes associated with the stimuli present when parental punishment of a forbidden response is administered. This fear of punishment aroused by the temptation to perform the forbidden behavior then acts as an acquired drive (20, 26) or conditioned negative reinforcer (7, 21, 25) to deter a child from performing the response even when the parent is not present to punish directly.

Much of the hypothetical formulation which guided the study grew out of findings of previous investigations in which childrearing practices and measures of guilt were associated (1, 2, 3, 8, 19, 22, 30). Since it was assumed that a child who had experienced guilt would later tend to resist temptation in order to avoid the unpleasantness of remorse, it was expected that the antecedents for guilt, found in these previous studies, and those for resistance to temptation would tend to be the same. These earlier findings (as well as other theoretical considerations) led to the expectation that severity of weaning and toilet training (30) and certain aspects of sex training (11, 24) should be related to resistance to temptation. It was expected, too, that associations found earlier of guilt with parental use of reasoning and love-oriented techniques of discipline (in contrast to arbitrary and physical punishment) would also be found between resistance to temptation and these parental disciplinary practices (1, 3, 8, 19, 22, 30). In addition, our theory of identification led us to expect that children of restrictive parents (as contrasted with permissive parents) would be more likely to engage in extensive covert role-practice of their parents' actions and statements, and thus would more thoroughly "internalize" parental values and be more likely to resist temptation.

Without reference to the concept of identification, it is possible to make predictions concerning the features of socialization which ought to be related to resistance to temptation, simply on the basis of reinforcement learning theory. First of all, consistency should be important. If the parents always interpret the same acts as being "wrong" and they employ methods (such as keeping close track of the child's whereabouts) for insuring that the child is almost always caught when deviating, the child's resistance to temptation should be increased. Presumably, both fairly high levels of punishment for deviation and reward for conformity should also be associated with high resistance to temptation. Furthermore, Solomon and Whiting[5] have suggested that temporal parameters involved in the application of punishment for deviations may be crucial in producing resistance to temptation versus guilt. According to this theory, the stimuli which have been present during the administration of punishment become, through conditioning, capable of arousing anxiety in anticipation of the actual punishment. If the parent has usually administered punishment *before* the child could commit the forbidden response, the cues producing anxiety should be maximum when the child is approaching the tabued acts, i.e., is being tempted, and thus should produce maximum resistance to temptation. On the other hand, punishment usually administered *after* the response has occurred should condition anxiety to the cues present at that time and thus be related to guilt or feelings of remorse after a deviation. Whiting's specific resistance to temptation hypothesis (27) adds to these conditions the prediction that punishment administered by a generally nurturant agent is more effective in producing resistance. There is some evidence from studies with dogs which partially support these hypotheses (6, 9, 27).

Some previous investigations related conditions of the home background with conformity to rules but were either very general (14) or obtained the background measures from the subject's own report (19). More recent research by Sears, Maccoby, and Levin (22) related many areas of childrearing practices to a rating of the child's current evidence of conscience, all measures based on mother interview data. The present study replicates much of this latter investigation and also provides a behavioral measure on the child to be used as an index of resistance to temptation.

In line with our theoretical interests and the final research design, 15 general areas of parent-child characteristics relevant to childrearing practices were considered: activity level of the child now and during infancy; weaning; toilet training; sex training; aggression training; *S*'s tendency to be emotionally upset; restrictions in infancy and currently; independence training; level of demands; clarity and consistency of demands, rules, and restrictions; mother's attitude toward *S*'s knowledge of games, rules, and cheating; techniques of discipline; use of positive reinforcement; mother's affectional warmth; and father's affectional warmth and level of demands. The general questions which this report considers in relation to these content areas are:

1. Do early childrearing practices—specifically, weaning, toilet training, and sex training—relate to conformity to rules in the four-year-old?

2. Are techniques of punishment (love-oriented as contrasted with physical punishment or object-oriented discipline) related to resistance to temptation?

---

[5] Personal communication, 1956.

3. Does the closeness of parental control relate to resistance to temptation? Two facets of this question deserve study: (a) parental consistency and (b) the general level of demands and restrictions.

4. Is resistance to temptation related to the timing of punishment (that is, whether punishment is usually administered just before, instead of after, the deviant act)?

5. Is resistance to temptation related to the identity of the disciplinary agent or to the degree of affectional warmth that characterizes this agent?

METHOD

Subjects

The 77 children, 40 boys and 37 girls, in this study were all four-year-olds and enrolled in private nursery schools in four districts of the greater Boston area.[6] The *S*s were tested between April and June of 1957, and the mothers were interviewed between May and August. In general, these *S*s came from well-established, middle-class homes. Only six fathers and 18 mothers had not attended college. Most of the fathers were business or professional men. The religious preferences were Jewish for 38 and Protestant for 30 of the families with the other nine families being mainly mixed in preference. The ethnic backgrounds of the parents were chiefly British, Northwestern, or Eastern European.

Procedure

The major dependent variable is a behavioral measure of whether or not the child conformed to rules of a game when he played it by himself. A bean bag game was built which was interesting to the *S*s, appeared to be controlled by the player's skill, permitted deviation from the rules easily, and yet was in fact completely controlled by an experimenter. The game consists of a 1 ft. by 4 ft. board on which there are five lights in a row. As presented to the *S* and as it appeared to the *S*, hitting a wire which was stretched along behind the board (but which he could not see from his throwing position) would turn on a light and ring a bell. The game was placed against a wall to make it seem plausible to the child that if a bag hit high on the wall it could fall down on the wire and make a score. Regardless of how many times *S* hit the wire with his bean bag, however, he got a score only when a hidden experimenter turned on a light. Individual differences in skill were thus controlled, as were schedules of reinforcements.

In order to test the children in a rather familiar setting as well as to avoid the costs of transporting them to the university laboratory, a portable observation booth was built. This booth and the game were installed in an empty room of the *Ss*' nursery school.

---

[6] Originally 85 *S*s were tested, but only 79 of the mothers were located for home interviews. Two of the obtained mother interviews were discarded due to failure of recording equipment.

Each *S* was shown how the game "worked," was given standardized practice in throwing the bags, and was carefully instructed in the rules. The rules were to throw each of the five bags only once over the board while standing on a foot marker placed approximately 5 feet from the front of this board.

After it was clear that *S* understood the rules, a second game was introduced.[7] After three minutes of the second game, *E* brought *S* back to the bean bag game and reviewed the rules. Then, a tray of toys, pretested for their attractiveness, was shown *S*, and he was asked to select one he would take were he to "get enough lights on to win a prize." This was done to be sure he focused on them enough to develop a strong desire to have one.

E then looked at his watch and said he had to make a phone call but that he wanted *S* to play the game according to the rules while he was out of the room. The implication was that *S* was to play for the prize. *S* was shown how to hook the door from the inside, so that no one would "bother" him while he was playing the game. *E* said he would knock when he wanted to be readmitted.

If *S* played according to the rules during the three-minute test, he received only one light (after his second throw) in contrast to the three lights "earned" in each practice session. Additional lights were given for each act of cheating: stepping forward, moving the foot marker, retrieving bags incorrectly, and hitting the wire with the hand. During this time, the hidden *E*, controlling the game, recorded *S*'s behavior.

When *E* returned, another game was played, "this time for the prize." This statement was intended to reduce any guilt that might have been aroused by cheating, to relieve feelings of failure for noncheating *Ss*, to avoid reinforcing cheating, and to have a check on whether *S* evinced understanding of the rules by playing according to them with *E* present. All *Ss* were given four lights in this last game and awarded a prize.

Before continuing, we would like to explain part of this procedure. It can be argued that many motives are aroused by the temptation test. Examples of motives that would lead to obtaining a high score through cheating are (a) to please *E* by throwing well, (b) to fulfill *S*'s own achivement strivings, competitive tendencies, and/or need to outwit authority, and (c) to win the toy. Motives that would promote resistance to temptation include (d) to please *E* by conforming to his rules, (e) to avoid external punishment (actual or imagined), and (f) to comply with *S*'s own internalized standards. Instead of attempting to control each possibility of these sorts, our decision was to test resistance to temptation under conditions which attempted to maximize all possible motives to deviate. Thus, intensity of temptation in toto would be controlled, and individual differences in temptation would be less readily attributable to mere variations in desire to transgress. To minimize risk and to maximize temptation, all windows through which *S* could be observed were covered with curtains, *E* acted pleased when *S* was given a "hit" during the practice games, the toys were shown and *S* was asked to choose the one most attractive to him, *S* was instructed to hook the door from the inside during the test, and each *S* was given the same reinforcement schedule of only one "hit" during the test if he conformed to the rules. In learning the rules, *S*s got three "hits" in each practice game. Thus, we felt confident that they considered getting just one light was a poor showing.

---

[7] Part of the full procedure irrelevant to this presentation will not be discussed here. The complete description is in Burton, Allinsmith, and Maccoby, "Resistance to temptation in relation to sex of child, sex of experimenter, and interruption of attention" (not yet published), which presents material on the effects of the experimental variables.

An interview schedule was developed to obtain information for rating those childrearing variables and indexes of early childhood behavior considered theoretically important. Most of the questions were open-ended, and the interviewers were trained to make appropriate probes in order to get sufficient details for scoring. All interviews were recorded on Audograph discs, and transcripts made later. The interview averaged two hours and was very much like that described in Sears, Maccoby, and Levin (22). There were, however, some variations in content.

Parent training measures. We sought indexes involving as few inferences as possible. Therefore, ratings scored in each of the content areas were generally based on specific information in the protocol, rather than on global judgments.

All interviews were scored by the senior author. Forty-four interviews were coded by at least two people; but, since some interviews had been used for training, only 24 were used for testing reliability. Of the 142 scales scored, 28 were discarded for not reaching a Pearsonian r of .70. The remaining scales had reliability coefficients ranging from .70 to 1.00, with a mean of .86.

Summary scales were made of the childrearing variables on the reasoning that each original scale could be considered analogous to an item in a test of a general area of child training. The method of computing these summary scales was first to choose the individual scales considered psychologically related, that is, as potentially measuring the same underlying training area. Then, tables of intercorrelations were made, and from these tables those scales found to have internal consistency were chosen. The criterion for internal consistency was that any rating to be included must have a majority of positive correlations with the other ratings and should not have any significant negative correlations with any of the ratings finally chosen. The scores of the original scales were then added. These sums were categorized into 5-point scales as close to normal distributions as possible for the final summary score.

Child measures. The resistance to temptation score was based on the child's conformity to the rules of the game during the temptation test. This was a 7-point scale based on a count of the number of bags *S* threw correctly before cheating, with categories at the extremes for immediate cheating and for no cheating during the test. Thus, "1" was for deviating before any bags were thrown; "2" through "6" were the scores obtained by adding 1 to the number of bags correctly thrown before first deviation; and "7" was for never deviating during the test period. This measure had almost perfect reliability.

An "evidence of conscience" measure was obtained from a rating of the mother's report in the interview, the same measure as in Sears, Maccoby, and Levin (22). This measure was made from the mother's report of how *S* behaved when he had done something wrong and before the mother knew about it. This rating had a scoring reliability coefficient of .83.

Results

The individual measures were divided at the median and fourfold tests of association made. Whenever possible, Fisher's exact probability tests were used; otherwise, χ^2 was computed, with all probabilities based on two-tail tests and corrected for continuity. The summary scales were correlated with both child measures using product-moment correlation.

Childrearing and Early Childhood Behavior Related to Resistance to Temptation

Table 1 summarizes the relations between the individual childrearing scales and the resistance to temptation measure. Also included are a few indexes of child reactions to socialization practices: the information about these reactions was also obtained in the mother interviews. Only those scales significantly associated with resistance to temptation or of special theoretical relevance are included.

Activity level. The three activity scales were highly intercorrelated, with correlations decreasing in magnitude as the age differences increased. Both infancy (before one year) and early childhood (between first and second year) activity levels were positively related to resistance to temptation for boys (with a corresponding but nonsignificant trend for girls). The current activity (age at the time of interview) rating, however, was not significantly related to conformity to rules, and the data are not presented.

Early childrearing. The only weaning variable which was related to temptation was an over-all rating of severity which was positively related (for boys only). The more severe the weaning, the more conformity to rules. This rating measured how gently the mother changed *S* from sucking to drinking, at the low end of the scales, and how abruptly and severely, at the other end. Age at beginning and completion of weaning were separate ratings and proved to be uncorrelated with resistance to temptation.

Time between initiation and completion of bowel training and the age when *S* stayed dry during the day were related to resistance, but again these relations were found only among boys: the longer bowel training took and the later the completion of daytime dryness training, the more conformity to rules. The daytime dryness variable was dichotomized at the median for each sex before being combined for the test of association. The median for girls was 24 months, for boys, 34 months. While it was also found that the age for beginning bowel training was significantly related to the amount of time it took for completion (the earlier training began, the longer it

TABLE I

CHILDREARING AND CHILD BEHAVIOR RELATED TO RESISTANCE TO TEMPTATION

		RESISTANCE TO TEMPTATION				
		BOYS		GIRLS		
		High	*Low*	*High*	*Low*	*Significance*
Activity level during first year	High	14	4	13	9	Boys* Girls Combined: $\chi^2 = 3.67$
	Low	8	13	7	7	
Activity level between one and two years	High	14	4	11	5	Boys* Girls Combined: $\chi^2 = 7.53^{**}$
	Low	8	14	9	12	
Severity of weaning (age of beginning not considered here)	High	14	5	11	9	Boys* Girls Combined: $\chi^2 = 2.55$
	Low	7	12	8	8	
Time between initiation and completion of bowel training	"Short"	5	13	7	7	Boys** Girls Combined: $\chi^2 = 4.00^{*}$
	"Long"	13	4	12	10	
Age when stayed dry during day	Early	8	15	7	10	Boys** Girls Combined: $\chi^2 = 7.85^{**}$
	Late	13	3	13	7	
Explanations given about origin of babies						
Rarely		10	15	13	8	Boys* Girls
Often (whenever *S* might ask)		11	3	7	9	
S's reaction when he learned about sex differences	High	6	9	3	9	Boys Girls* Combined: $\chi^2 = 6.46^{*}$
	Low	16	9	16	7	
S's reaction to being sent to room	High	9	10	8	7	
	Low	13	7	12	9	
Timidity	High	7	5	9	7	
	Moderate	7	9	6	4	
	Low	8	4	5	6	
Techniques used for teaching S not to touch objects						
Stops before touches		9	12	7	11	Boys Girls Combined: $\chi^2 = 4.37^{*}$
Waits to see if he really plays with it		12	6	13	6	

(*table continued on next page*)

TABLE 1 (*continued*)

CHILDREARING AND CHILD BEHAVIOR RELATED TO RESISTANCE TO TEMPTATION

		RESISTANCE TO TEMPTATION				
		Boys		Girls		
		High	*Low*	*High*	*Low*	*Significance*
Mother's attitude toward cleanliness	Relaxed	13	2	4	5	Boys** Girls
	Concerned	9	16	16	12	
Does S understand what cheating means?	Yes	10	7	5	9	
	No	12	11	15	7	
Usual technique of discipline						
Reasoning		0	0	1	1	
Visible suffering		0	0	0	1	
Withdrawal of love		0	0	1	0	
Isolation		7	8	6	5	
Deprivation of privilege or object		3	4	0	4	
Scolding		2	2	7	3	
Physical punishment		10	4	4	3	
Scolding and physical punishment vs. all others: Boys Girls Combined: $\chi^2 = 2.79$ ($p < .10$)						
Frequency of Mother's use of reasoning as teaching technique	High	8	12	5	9	Boys Girls Combined: $\chi^2 = 5.41$*
	Low	14	6	15	8	
Use of "I'm not going to talk to you until you behave better"	Yes	12	3	5	6	Boys* Girls
	No	10	15	15	11	
Use of deprivation of privileges	High	10	8	8	13	Boys Girls*
	Low	10	10	12	3	
Frequency of Mother's playing with S just for her own pleasure	High	14	8	10	15	Boys Girls*
	Low	8	10	9	2	
S's initial anxiety in the experiment	High	5	9	8	1	Boys Girls*
	Moderate	8	7	8	9	
	Low	8	3	4	7	

NOTE.—Throughout this table the variable being considered was divided at the median.
* $p \leq .05$. ** $p \leq .01$.

took), it is nevertheless true that the age of beginning bowel training was not in itself related to resistance to temptation. Furthermore, our rating on severity of toilet training did not predict resistance to temptation.

The only rating of sex training to be significantly related to resistance to temptation concerned explanations about origin of babies. Parental offering of such explanations in response to a child's query was associated with resistance for boys. The trend is opposite for girls, though nonsignificant. Although we had reliable scales on the degree of sex permissiveness and degree of mother's sex anxiety, they did not correlate with the children's resistance to temptation. The original reason for including these measures of sex training was based on the hypothesis that children with high sex anxiety, whatever the contributing factors, would tend toward an inability to delay gratification which in turn would incline the child to cheat in the temptation test. In general, then, the conditions which ought to produce sex anxiety in the child—namely, severe, nonpermissive training in regard to sex—did not prove to be associated with yielding to temptation, except for the finding about giving information to boys. Another significant association interpreted as supporting this hypothesis is that between cheating and a high reaction when *S* first found out about sex differences (especially for girls), such as a great amount of curiosity or feeling that something was wrong with himself or herself.

Anxiety. In addition to anxiety about sex, it was hypothesized that in general a highly "anxious" child would be low on self-control and delay of gratification with an increased tendency to cheat. The indexes of anxiety in *S* based on the interview were ratings of his general timidity in meeting people, amount of upset over making mistakes, regression to wanting a bottle or to wetting the bed, amount of upset when sent to his room, as well as intensity of reaction to sex differences. The only "anxiety" rating from the interview which was related to resistance to temptation was the high reaction to sex differences associated with cheating. Table 1 shows the distributions on the interview ratings of *S*'s reaction to isolation and to meeting strange people. The other ratings of *S*'s emotionality were evenly distributed. In addition to the ratings based on the interview, there were two ratings of the child's "anxiety" during the experimental situation. An "initial anxiety" rating was made from *S*'s behavior upon being brought into the experimental room and was based on the amount of hesitancy he showed when leaving the classroom, fearfulness of *E* and the bean bag game, and reluctance to permit the teacher to leave him when coming into the experimental room. The other "anxiety" rating was based on *S*'s reaction to *E*'s withdrawal of attention during the experimental situation. The latter showed no trends, but the relation between the behavioral rating of initial anxiety and resistance to temptation is shown in Table 1. It is seen that for boys there is some tendency in the distributions for a high "anxiety" rating to be associated with cheating. The distributions for girls suggest that two types of "anxiety" have been measured. One type seems to reflect the emotional instability which it was originally hypothesized would be related to cheating; the other type suggests a fearfulness of responding. These findings will be discussed further below.

Measures of aggression training had very low reliabilities and therefore did not permit a test of the relationship between this area of child training and cheating.

Timing of punishment. It was originally thought that the measure of the parent's timing in trying to stop *S* from touching something would be related to the ratings on restrictions, but it was not. This scale was, however, related to resistance to temptation. The categories for the scale ranged from stopping *S* before he touched the object ("1") to permitting him to touch it and then telling him not to touch it any more ("5"). The association indicated that letting a child actually touch objects before saying "no, no" is related to resistance.

Parental control. Only one of the 25 ratings measuring restrictions, demands, and autonomy was significantly related to resistance. This rating was the mother's attitude toward cleanliness, and the association was only for boys; the greater the mother's concern about cleanliness, the greater her son's tendency to cheat.

None of the scales attempting to measure the mother's consistency and clarity of rules, demands, and restrictions discriminated between high and low resistance. This was also true for the ratings of the mother's attitude about rules. For these last-named ratings, the mother was asked how often *S* played games with rules, whether he understood about rules and cheating, and how often she consciously taught him about cheating. It seems of interest to note that a mother's opinion of her son's knowledge of rules had little relation to his behavior; in contrast, girls whose mothers felt they definitely understood about following rules and about cheating tended to cheat. This tendency for girls which is nonsignificant when the Resistance to Temptation (RTT) scale is divided at the median becomes significant if "noncheaters" are compared with "cheaters" (Table 2).

TABLE 2

UNDERSTANDING OF CHEATING RELATED TO RESISTANCE TO TEMPTATION

	BOYS		GIRLS**	
	Noncheaters	*Cheaters*	*Noncheaters*	*Cheaters*
Does S understand about cheating?				
Yes	7	10	0	14
No	8	15	12	10

** $p \leq .01$.

Techniques of punishment. In regard to disciplinary techniques, physical punishment included such acts as spanking, slapping and shaking. Scolding was verbal reprimand, usually with a raised voice. The mother's taking away a toy or TV was scored deprivation of object or privilege. Isolation was primarily scored for sending *S* to his room. The prototypical response for withdrawal of love would be, "I can't love you when you do that." Visible suffering (4) was represented by the mother's saying to *S*, "You must not love me to act that way." Some mothers claimed to discipline their children by discussing why they wanted them to act in certain ways and expressed the attitude that they didn't believe in "punishment." The major disciplinary technique for these mothers was "reasoning." A rating of the mother's usual choice of discipline indicated a suggestive trend that scolding and physical punishment tend to be related to conformity when constrasted with all other disciplinary techniques (p between .10 and .05).[8] The other measures of physical punishment and scolding, such as frequency and intensity, did not show any strong trends of association with resistance. A consistent trend was found for high use of reasoning to be related to cheating. This tendency was found in the scales comparing preferred types of discipline and mother's belief in reasoning and was significant in the measure of the extent the mother used reasoning. For withdrawal-of-love techniques, only high use of not talking to *S* until he behaved was related to resistance, but only for boys. The only other negatively centered training technique which showed a significant association with temptation behavior was the high use of deprivation of privileges which was related to cheating for girls only.

The measures for the positively focused training techniques did not show any significant associations with behavior under temptation (and are not shown here), but the trends in the distributions were rather consistent. For boys the trend was for high use of positive reinforcement, such as praise, affection, tangible rewards, and privileges, to be related to resistance. The pattern for girls was for the use of privileges and tangible rewards, labeled "object-oriented techniques" by Sears, Maccoby, and Levin (22), to be related to resistance, but for high use of love-oriented practices of praise and affectional reinforcement to go with cheating.

Identity and warmth of disciplinary agent. The measures of the mother's warmth asked about her attitude toward *S*, what she liked about him, how demonstrative she was toward him in showing affection, how often she played with him for her own pleasure, and whether she took time to answer his questions. The only significant relation for these measures was for girls: those whose mothers said they frequently played with their daughters just for pleasure tended to cheat. For boys, the consistent trend, though not significant statistically, was for warm mothers to have sons who tended not to cheat. The same nonsignificant trends for girls and boys were found in the measures of father warmth; high warmth tended to be related to cheating in girls and conformity in boys.

Resistance to Temptation and Evidence of Conscience (Guilt)

In most theoretical formulations based on the concept of identification, it has usually been assumed that resistance and guilt would be positively related. The reward and punishment theory, however, led to the hypothesis that conditions for producing resistance to temptation and for producing guilt might be different, namely, administration of punishment before the act versus after the act, respectively. A test of association between these two "superego" indexes, resistance to temptation in the game and evidence of conscience, indicates that there tends to be a negative correlation between them. If all "cheaters" are compared with *Ss* who never deviated from the rules during the entire three-minute test period, this negative association is less than the .10 level of probability.

---

[8] The senior author made an over-all rating of the mother's preferred mode of discipline —psychological, reasoning, scolding, or corporal—which indicated corporal punishment was related to conformity for boys and girls. However, a check showed the rating to be too unreliable for any confidence. We would like to express our appreciation to Mrs. Faye Goldberg and Dr. John D. Campbell for performing this reliability check.

TABLE 3

EVIDENCE OF CONSCIENCE RELATED TO RESISTANCE TO TEMPTATION

	BOYS		GIRLS	
	Noncheaters	*Cheaters*	*Noncheaters*	*Cheaters*
Evidence of conscience (Guilt)				
High	8	18	4	16
Low	7	7	8	9

Sexes combined: $\chi^2 = 3.12$ ($p < .10$)

Childrearing and Evidence of Conscience (Guilt)

Sears, Maccoby, and Levin (22) have presented a theoretical formulation for the development of conscience. They reported correlations between childrearing practices and evidence of conscience (guilt) in the child which, although small and not all significant, were in support of their theory. As a check on the stability of these findings, the identical measure of conscience was obtained in this study and has been correlated with the summary scales for the childrearing measures. The resulting correlations were at least of the same magnitude as those previously reported and were generally higher. Low use of physical punishment (—.30),[9] high use of reasoning (.18), high use of praise as reward (.25), and high maternal warmth (i.e., affectional demonstrativeness) (.22) were significantly correlated with a rating of high conscience in both studies. Also considered theoretically important but not found to be significantly correlated in the previous Sears *et al.* study, the following practices were significantly associated in the present investigation: low use of tangible rewards and deprivation of privileges (—.16) and high use of isolation (.15).

In the present study the mother was asked several questions about her attitude toward the child's knowing about and following rules. The summary scale for these ratings was correlated .34 with the conscience rating, the highest of all the summary scales. This positive correlation indicates that the mothers who felt it was important for their child to learn about and to observe rules also felt their child evidenced many signs of a well-developed conscience. This somewhat severe attitude regarding conformity to rules being related to conscience contrasts to some extent with the "permissive" attitude expressed by the other correlations.[10] Considered together, however, these relations do suggest an internal consistency of childrearing beliefs shared by the mothers of this study. This consideration becomes more cogent in the light of the general tendency of these scales to correlate in the opposite direction (although not significantly) with our measure of resistance to temptation. This will be discussed further.

---

[9] Numbers in parentheses are product-moment correlations between the summary scale and the conscience rating in this study and are significant at or beyond the .05 level.

Discussion

An unexpected finding was the positive relation between the child's activity level (as reported by his mother) and his ability to resist temptation. We can only speculate about the reasons for this relationship. Two possibilities come to mind: the first is that active children are those in whom the parents have *permitted* activity, whose parents have not emphasized being quiet but have allowed the child to explore the environment fully. Perhaps, children treated in this way are simply more mature in all respects, including the kind of ego organization that makes possible conformity to rules. But this explanation does not seem entirely consistent with the fact that it was the child's level of activity *up to the age of two*, not after, that predicted resistance to temptation. We suggest an alternative explanation: parents who consider their child more active, either because the child is innately more active or because of the parents' own attitude toward how active a child should be, will feel that their child more often exposes himself to dangerous or forbidden things; these parents are therefore more often stimulated (by the actions of the child) into controlling actions designed to protect the child or the family's property; and, therefore, children who are considered highly active during the reaching-and-touching stage receive more training in *not* touching, a condition which facilitates the child's later ability to resist temptation in the absence of a controlling agent.

Our findings on the relation between early childhood socialization practices in the areas of weaning, toilet training, and sex training are fragmentary and not fully consistent. Severe weaning (usually meaning abrupt weaning in which the mother refuses to comply with the child's demands for the return of the bottle) was associated with high resistance to temptation in boys, as was *long-continued* bowel training, again for boys only. In the latter case it is notable that an early start of bowel training in itself was not significantly associated with resistance to temptation, nor was late completion of training. What is the special meaning of long-continued training? Often it means that a mother begins relatively early but for some reason does not succeed. Long-continued training frequently means that toilet training has become a battleground between the mother and child. The mother applies her demands unskillfully, with poor timing, and the child goes through periods of negativism and regressive behavior. Long-continued training appears to mean a greater total amount of pressure in this behavior-system than brief training does, be it early or late. We may have some basis, then, for the generalization that heavy socialization pressure on boys (as exemplified by severe weaning and extended toilet training) leads to greater resistance to temptation at a later time. But this interpretation must be tentative, considering that our scale on *severity* of toilet training did not correlate with resistance to temptation, nor did the amount of socialization pressure applied by the mother in the area of sex training.

---

[10] In the present study low severity of toilet training (−.22) and low activity (−.16) were also correlated with high conscience.

For boys the indexes of anxiety based both on the interview and on the behavioral ratings during the temptation test consistently showed trends which suggest that a high emotional reactivity is related to cheating. In girls there seemed to be two kinds of anxiety measured. One type, indicated by fearfulness of new situations and by timidity when meeting new people, tended to be associated with noncheating. This kind of anxiety seemed to predispose the child to feelings of fear during the test. The other type of anxiety seems to reflect a more general emotional instability and inability to delay gratification. This form of anxiety was suggested by the measures of *S*'s reaction to sex differences, to isolation, and to withdrawal of attention, and was associated with cheating. However, these interpretations are rather speculative since they are based as much on trends as on significant findings.[11]

In our introductory section, we discussed the possible outcomes of different timing of the mother's socializing actions. We mentioned the hypothesis that stopping a child *before* the execution of the forbidden act might lead to resistance to temptation, while stopping him after the action has actually begun might lead to guilt over deviation. Our finding does not confirm this prediction. The question we used in the mother interview was:

> Suppose you wanted him to learn not to play with something—like the TV set or matches. If you saw that he was tempted to touch it, would you stop him before he touched it or wait to see if he really played with it and *then* correct him?

The answers to this question do not provide us with exact information concerning the mother's timing of discipline. "Stopping before he touches it" covers a range from diverting the child to some other activity before he gets very close to the forbidden object to waiting until he actually starts putting out his hand and slapping the hand in the split-second before it touches the object. And similarly, "waiting till he really plays with it" could include applying discipline in the moment just following the child's first touching the object to letting him play with it for an extended period before stopping him and scolding or punishing him. We suspect that, since the objects asked about are either dangerous (matches) or quite expensive to repair if damaged (TV set), very few of these mothers actually allowed the child to play with the object to satiation. We suspect that the meaning of "wait till he actually plays with it" was more a matter of the mother's wanting to make sure that her intuition about his intentions was correct before applying discipline and that the discipline usually followed fairly closely upon the onset of actual touching. If this interpretation is correct, our finding becomes reasonable from the standpoint of learning theory, for it means that the correlation between discipline and the relevant cues for this particular action was very close in the "wait till he touches" condition. If the mother applies discipline when she only *thinks* the child is about to deviate but does not wait to make sure, she will sometimes be right and sometimes wrong, and the result will be poor correlation between the cues the child is actually responding to and the consequences of the action—an inefficient condition for learning. It is clear, however, that we need data on the actual timing of positive and negative reinforcement, recording much finer gradations in time intervals than can be obtained from interview questions. We expect theoretically that applying the socializing action *immediately* before or during the first phase of the subject's forbidden action would be more effective than applying it either earlier or later. But our data do not permit us to determine whether this is so.

---

[11] These two hypothesized kinds of anxiety seem to be similar to the factors found by Bell (5) to discriminate between neonates; the "tactile sensitivity" factor which is based essentially on the baby's reaction to aversive stimuli is analogous to fearfulness of new places, strange people, pain, etc.; the "arousal" factor seems similar to what we have considered emotional upset or instability.

The extent and severity of rules, restrictions, and demands, and the mother's consistency and clarity in establishing them would seem to be extremely important in establishing standards of behavior in *S*. However, the ratings which attempted to measure these variables showed few relations with response to temptation. The dearth of relations does not mean these variables have no effect on resistance; but it would seem to suggest that these variables, if indeed they *are* important, may have to be measured by tools other than the interview. Or perhaps the difficulty lies in our definition of what constitutes demands and restrictions.

It is difficult to account for the finding that girls who cheated tended to have mothers who felt their daughters understood about rules and cheating, but boys whose mothers said they understood about rules tended slightly to conform to them during the test. It may be there is a difference in the mother's attitude about what is referred to by "rules of the game" depending on the sex of her child. Games involving physical power and skill (games similar to our bean bag game) may be her referent for boys, whereas games more unlike our temptation test are associated with games played by girls. It is also possible that when mothers decide to teach their children to conform to rules they are more effective teachers for their sons than with their daughters, a possibility suggested by the Gewirtz and Baer (10) findings that social reinforcement is more effective when administered by an opposite-sexed experimenter.[12]

The indication that physical punishment has a tendency to be associated with resistance as compared with psychological punishment and use of reasoning would seem to be at odds with the implications from findings in previous reports (2, 3, 19, 22). Although the association between these variables reached a significance level just slightly under .10, the trend seems especially noteworthy because of its theoretical significance and its marked opposition to the expectation from these previous conclusions. These discrepancies may be due to poor measurement, differences in dependent measures, differences in ages of *S*s, and/or sources of measures. In Allinsmith's (2, 3) studies the measures of guilt and resistance to temptation were based on story completions and employed older *S*s than used in this study. MacKinnon (19) obtained his measures of parental discipline from a retrospective account of the *S*s who were also older. Sears, Maccoby, and Levin (22) had the same age *S*s but based both antecedent and consequent measures on the mother interviews, with the consequent measure being of guilt rather than of resistance to temptation. In contrast to the conclusions of these previous investigations, the trends in the present data suggest that scolding and physical punishment may be more efficient in producing resistance to temptation in four-year-olds than is psychological punishment, and certainly more effectual than attempting to train by just "reasoning." This evidence also suggests that, at least at this age, just understanding what an adult wants him to do or fearing psychological punishment is not as effective in deterring deviation by a child as is fearing physical punishment.

Our results then suggest that there is no simple relation between "permissive" or "severe" childrearing practices and resistance to temptation. Some training techniques which we found associated with conformity to the rules seem to reflect severe socialization practices. This severity grouping is suggested by the associations between the behavioral conformity and the interview measures of high activity, severity of weaning, long toilet training, use of scolding and physical punishment, and low use of "reasoning" as a disciplinary technique. On the other hand, resistance was also associated, for boys only, with explanations about sex, permissive attitude toward cleanliness, and not talking with him when he misbehaved (psychological punishment), and, for both boys and girls, with permission to touch an object before saying "no." This latter "cluster" of training practices suggests an emphasis on appealing to the child's cognitive awareness of what the parent wants. These groupings indicate the possibility that the differences in results between this study and the previous studies, regarding psychological and physical punishment, may in large part be due to the differences in the ages of subjects: previous studies have used older *Ss*. The hypothesis is suggested, here, that the more direct, severe, physical modes of training are most efficient and effective at the early ages of development. But with greater cognitive development, and perhaps increasing identification, the psychological techniques gain in power to influence the child's internalization of parental standards. The argument is that the direct forms of training represent a procedure similar to conditioning employing an unconditionally aversive stimulus; whereas at a later age the child is able to "comprehend" the more symbolic, psychological forms of punishment which may have acquired their aversive value.[13] This hypothesis of cognitive development also might account for the ineffectiveness of "reasoning" as a disciplinary technique for these four-year-olds.[14]

---

[12] For a somewhat different conclusion from other data, see Gump and Kounin (13). These authors conclude that boys are behind girls in the development of school-related value orientations due to developmental lag, identification difficulty, and/or intensification of preschool sex differences. This interpretation seems to correspond to that of Sears, Maccoby, and Levin (22) that boys lagged behind girls in the development of conscience.

The trend that boys from homes with high parental warmth tended to resist temptation, although not significant, is consonant with the results of the previous study (22). The direction of this result also supports Whiting's hypothesis that the administration of sanctions by a normally nurturant agent will be more effective (27). In contrast, a "cold" home was associated with conformity to rules for the girls. Sears, Whiting, Nowlis, and Sears (23) and Hartup (15) have advanced an identification hypothesis to account for this differential effect. Their argument is that girls are more sensitive to punishment by the mother because the mother is the girl's identificand. The finding which shows high use of withdrawal of love by the mother related to conformity for boys and not for girls in the present study tends to be contrary to this hypothesis. However, the findings that mothers in general are more nurturant to girls, especially during infancy, and use "psychological" forms of punishment more frequently with girls and physical punishment more frequently with boys (22, 23) are background conditions for considering these two hypotheses and indicate that they are not in opposition.[15, 16]

---

[13] It seems plausible that psychological discipline may become even more "painful" to *S* than corporal methods once he learns to respond to it, in that, if the parent slaps him, the punishment is over; but the mother's not talking to him for half an hour lasts longer.

[14] An investigation by Robert E. Grinder using 11-year-old subjects gives some support to this hypothesis (12).

[15] It has been reported that mothers of girls "reverse their policies, in terms of severity of frustration, from the infancy years to the later preschool years" (i.e., from high warmth and nurturance to a relatively greater severity of demands) and are "more consistent with respect to this variable" for boys (23). This evidence would suggest that the currently "cold" mothers of girls may have been "warmer" during the infancy years—a condition which brings together the hypothesis regarding the greater effectiveness of a basically nurturant agent with the identification hypothesis for our data.

[16] It is possible that innate constitutional differences between the sexes operate in regard to sensitivity to physical punishment. There is some evidence that the pain receptors of girls have a lower threshold than those for boys (5, 16). This greater sensitivity to pain would produce more fear associated with the cues of punishment and should lead to increased resistance to temptation for girls relative to boys under equal amounts of physical punishment. Again, this hypothesis should be considered in light of the evidence that boys

The negative relation between the measure of resistance to temptation and the rating of conscience lends support to the hypothesis that the conditions for guilt and resistance to temptation are different. This hypothesis that moral development is a multiple rather than a single process was offered by Allinsmith (1, 2) as an interpretation of his findings and has been discussed by Maccoby (17) in a paper which utilizes additional data from recent and still unpublished studies. It is also possible that a greater number of significant relations which we found between the childrearing variables and the conscience measure, as compared to the temptation measure, indicates a "halo" effect arising from the fact that the conscience measure and the childrearing variables are both taken from the mother interview. Furthermore, the mother's perception of the child cannot be independent of either her socialization practices or her belief about their outcome. Although the validity of childrearing measures based on interviews is always difficult to assess, the number of associations found due to "halo" effect would probably be lessened when the consequent variable is a systematically quantified measure based on the child's actual behavior as rated by a neutral observer.

SUMMARY

In this study selected childrearing techniques and child behaviors as rated from a mother interview have been related to an experimental measure of resistance to temptation in four-year-old children.

The reported findings and discussion indicate that conformity to rules is a highly complex domain for psychological investigation. Resistance to temptation appears to be related to a number of factors, some of which have been tapped in this study and probably others which will be revealed in the future.

Severity–Permissiveness

There appears to be no simple relation between resistance and the mother's attitude toward severe and permissive training practices. Our results indicate that some measures which seem to reflect heavy socialization demands are more effectual than permissive practices. These measures pertain to activity level of the child during infancy and early childhood, severity of weaning, long toilet training, use of scolding and physical punishment, and infrequent use of "reasoning" for discipline. On the other hand, another set of variables found significantly related to resistance (primarily for boys) appears to indicate a permissive attitude. These practices were explanations about sex, lenient attitude toward cleanliness, and use of withdrawal of love as punishment.

---

are physically punished more than girls (22, 23). Directly related to resistance to temptation are the results of the dog experiments which indicated that constitutional factors of breed and sex, seemingly related to sensitivity to punishment, were important in determining resistance to temptation (6, 9). Unfortunately, the data from the present study do not permit direct consideration of this hypothesis.

Age and Cognitive Development

An inspection of these two groupings of variables suggested that the first set, which indicates severity, is of greatest relevance during the early period of the child's development. The permissive cluster seems more relevant to a later period in development and seems to require a greater cognitive awareness than would be necessary for the severe training techniques to be effective. These two groupings and the differences between the findings of this study and those of previous reports led to a hypothesis regarding differential child training techniques and cognitive development. This hypothesis is that the more severe, direct, physical forms of childrearing are most efficient during the child's early years and that with increasing cognitive development the more symbolic forms of socialization techniques gain in effectiveness. The generality of this hypothesis is tentative, however, in that the permissive cluster consists of variables related to resistance primarily for boys only. Furthermore, the interpretation regarding differential importance for developmental ages is inferential and has no direct support from this study.

Sex Differences

Readiness to answer questions about sex, lenient attitude toward cleanliness, and use of withdrawal of love were associated significantly with conformity for boys and nonsignificantly with cheating for girls. A home atmosphere of high affectional warmth was related significantly with cheating for girls and nonsignificantly with conformity for boys. Two hypotheses were discussed relevant to these sex differences. Firstly, Whiting's hypothesis that a normally nurturant agent is more effective when administering sanctions is appropriate for boys. Secondly, the association for girls between a "cold" home atmosphere and resistance suggested an identification hypothesis, with girls being more sensitive to their mother's punishment than are boys.

Temptation and Guilt

The associations found by Sears, Maccoby, and Levin (22) have been supported in regard to mother interview measures of the child's evidence of conscience or guilt and of childrearing antecedents. However, the behavioral measure of resistance to temptation tended to be associated with opposite sets of antecedents and to be negatively related to the mother interview measure of conscience. These findings indicate that resistance to temptation and guilt should be studied as distinct phenomena and should not be considered as simply alternative measures of internalized standards.

REFERENCES

1. ALLINSMITH, W. The learning of moral standards. Unpublished doctoral dissertation, Univer. of Michigan, 1954.

2. Allinsmith, W. The learning of moral standards. In D. R. Miller & G. E. Swanson, *Inner conflict and defense.* Holt, 1960. Pp. 141-176.

3. Allinsmith, W., & Greening, T. C. Guilt over anger as predicted from parental discipline: a study of superego development. *Amer. Psychologist,* 1955, 10, 320. (Abstract)

4. Baldwin, A. L. *Behavior and development in childhood.* Dryden Press, 1955.

5. Bell, R. Q. Relations between behavior manifestations in the human neonate. *Child Develpm.,* 1960, 31, 463-477.

6. Black, A., Solomon, R. L., & Whiting, J. W. M. Resistance to temptation as a function of antecedent dependency relationships in puppies. *Amer. Psychologist,* 1954, 9, 579. (Abstract)

7. Dinsmoor, J. A. Punishment: I. The avoidance hypothesis. *Psychol. Rev.,* 1954, 61, 34-46.

8. Faigin, H. Child rearing in the Rimrock community with special reference to the development of guilt. Unpublished doctoral dissertation, Radcliffe Coll., 1953.

9. Freedman, D. G. The effects of indulgent and disciplinary rearing in four breeds of dogs. *Science,* 1958, 127, 585-586.

10. Gewittz, J. L., & Baer, D. M. The effect of brief social deprivation on behaviors for a social reinforcer. *J. abnorm. soc. Psychol.,* 1958, 56, 49-56.

11. Goethals, G. W. A study of the relationship between family esteem patterns and identification, the internalization of values, and aggression of a group of four-year-old children. Unpublished doctoral dissertation, Harvard Univer., 1955.

12. Grinder, R. E. Behavior in a temptation situation and its relation to certain aspects of socialization. Unpublished doctoral dissertation, Harvard Univer., 1960.

13. Gump, P. V., & Kounin, J. S. Milieu influences in children's concepts of misconduct. *Child Develpm.,* 1961, 32, 711-720.

14. Hartshorne, H., & May, M. A. *Studies in deceit.* Macmillan, 1928.

15. Hartup, W. W. Nurturance and nurturance-withdrawal in relation to the dependency behavior of preschool children. Unpublished doctoral dissertation, Harvard Univer., 1955.

16. Lipsitt, L. P., & Levy, N. Electrotactual threshold in the neonate. *Child Develpm.,* 1959, 30, 547-554.

17. Maccoby, E. E. The generality of moral behavior. *Amer. Psychologist,* 1959, 14, 358. (Abstract)

18. Maccoby, E. E. Role-taking in childhood and its consequences for social learning. *Child Develpm.,* 1959, 30, 239-252.

19. MACKINNON, D. W. Violation of prohibitions. In H. W. Murray (Ed.), *Explorations in personality*. Oxford Univer. Press, 1938. Pp. 491-501.

20. MILLER, N. E. Learnable drives and reward. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. Wiley, 1951. Pp. 435-472.

21. SCHOENFELD, W. N. An experimental approach to anxiety, escape and avoidance behavior. In P. H. Hoch & J. Zubin (Eds.), *Anxiety*. Grune & Stratton, 1950. Pp. 70-99.

22. SEARS, R. R., MACCOBY, E. E., & LEVIN, H. *Patterns of child rearing*. Row, Peterson, 1957.

23. SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, P. S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, 47, 135-203.

24. SHAPLIN, J. T. Child training and the identification of pre-adolescent boys with their parents. Unpublished doctoral dissertation. Harvard Univer., 1954.

25. SKINNER, B. F. *Science and human behavior*. Macmillan, 1953.

26. SOLOMON, R. L., & BRUSH, E. S. Experimentally derived conceptions of anxiety and aversion. In M. R. Jones (Ed.), *Nebraska symposium on motivation*, 1956. Univer. of Nebraska Press, 1956. Pp. 236-248.

27. WHITING, J. W. M. Fourth presentation. In J. M. Tanner & Barbel Inhelder (Eds.), *Discussions on child development: II*. Tavistock Publications, 1954.

28. WHITING, J. W. M. Sorcery, sin and the super-ego: a cross-cultural study of some mechanisms of social control. In M. R. Jones (Ed.), *Nebraska symposium on motivation*, 1959. Univer. of Nebraska Press, 1959. Pp. 174-195.

29. WHITING, J. W. M. Resource mediation and learning by identification. In I. Iscoe & H. W. Stevenson (Eds.), *Personality development in children*. Univer. of Texas Press, 1960. Pp. 112-126.

30. WHITING, J. W. M., & CHILD, I. L. *Child training and personality: a cross-cultural study*. Yale Univer. Press, 1953.

MILIEU INFLUENCES IN CHILDREN'S CONCEPTS OF MISCONDUCT[1]

PAUL V. GUMP *and* JACOB S. KOUNIN*

Wayne State University

One task of the school is to establish a set of mores or values which will enable it to function effectively. Introduction of this set of school values occurs early in the child's school experience. The investigation to be reported here studied differences in children's ideas of misconduct as these were related to the milieus (school vs. home) in which misconduct could occur. In order to test generality of findings across sex and age variables, boys and girls were interviewed during their first semester in the first grade and again two years later. As a secondary part of the study, the influences of sex and age were also studied.

METHOD

Subjects were 113 boys and 114 girls from six urban schools representing a range of socioeconomic backgrounds. After a brief getting acquainted exchange, children were asked the following questions in the first and, later, in the third grade:

1. What is the worst thing a child could do at school?
2. Why is that so bad?
3. What is the worst thing a child could do at home?
4. Why is that so bad?

---

* College of Education, Wayne State University, Detroit 2, Michigan.

[1] Financial support for this study was received from the College of Education, Wayne State University, and from the National Institute of Mental Health, United States Public Health Service. We are indebted to the personnel of the Detroit Public Schools for making facilities and subjects available. Dr. Bruce Biddle made significant contributions to code development.

Child Develpm., 1961, 32, 711-720.

A child's answers to these questions can be examined in terms of the kind of misconduct mentioned and the type of explanation given for why the misconduct was bad. Responses of the children were subjected to a comprehensive coding schedule for which a satisfactory reliability had been established; median intercoder agreement for the scales and categories used was 90 per cent. For sake of clarity, codes will be described when results from their use are reported.

RESULTS

Influence of Milieu in Misconduct Type

Misconducts were investigated from three points of view: their general category, their specific type, and their moral seriousness.

Five major categories were sufficient to describe the general categories of misconduct:

1. Violations of rules—no person is directly harmed. "Go visit a friend without asking first." "Talk when you're supposed to study."
2. Assaults on other children. "Pull my sister's hair." "Punch a kid in line."
3. Breaking or damaging property. "Break the kitchen window." "Write on the desks."
4. Nonconformity with adult leader of the setting. "Not mind their mothers." "Don't do what teacher says."
5. Miscellaneous—This includes misconducts which were mentioned infrequently and were too disparate to be combined into one meaningful group. "Steal." "Sneak out." "Tease somebody."

The results of coding children's responses into general categories are displayed in Table 1.

The most consistent difference between school and home on category of misconduct was that related to *breaking*; this category was always most frequent for the home milieu. (When sexes were combined and home and school were compared, p values were .01 for both the first and third grade administrations.)[2] *Nonconformance* tended also to be a home sin but this was most pronounced for the first grade ($p < .01$); the same trend continued for third grade girls ($p < .08$). It is clear that direct disobedience or defiance of teachers was a rarely reported misconduct, but that such affronts to parents were reasonably frequent. A misconduct reported most frequently for school was *assaults* on children. Boys always had more assaults for school than for home ($p < .01$); girls showed this same school-home contrast in the third grade ($p < .02$). The failure of the girls to show more assaults for school than for home in the first grade was related to their heavier use of *violations* for school rather than home misconducts ($p < .001$). This fact, plus other data to be presented, indicated

---

[2] Differences between settings and between ages are those involving the change of a person's response from one situation to another; for these comparisons the McNemar test was used (2). Differences between sexes were tested with chi square.

TABLE 1

GENERAL CATEGORIES OF MISCONDUCT REPORTED BY CHILDREN

	Violations of Rules	*Assaults on Children*	*Breaking Property*	*Nonconform with Adult*	*Miscellaneous*
First Grade Boys					
Home	38	20	26	11	18
School	46	38	10	2	17
First Grade Girls					
Home	26	22	22	26	18
School	65	21	3	5	20
Third Grade Boys					
Home	50	16	25	3	19
School	41	30	16	4	22
Third Grade Girls					
Home	38	11	27	15	23
School	50	29	12	8	15

that girls' reactions to the impact of school in the first grade differed from boys' reactions.

The qualities of a misconduct are not fully described by the above general category system. Specific misconducts appeared with considerable frequency and seemed highly related to milieu, sex, and age variations. The large use of violations by girls for school derived from frequent reporting of *talking* as a misconduct. Talking was reported by one girl in four in the first grade. By the third grade this had been reduced to one in eight. *Fire play* was another specific misconduct, which showed marked differences between home and school; this was predominantly a boy misconduct, but girls also reported it. Fire play was more frequently a home misconduct for boys and for girls in the first grade ($p < .001$ and .01, respectively) and for boys again in the third grade ($p < .001$). By the third grade one boy in four reported fire misconducts for home—one in 16 for school.

A final aspect of the kind of misconduct related to moral seriousness. The same category may include both morally trivial and morally reprehensible acts. For example, assaults on children can include both: "Sock a kid in the arm and run" and "Beat up little kids who didn't do nothin' to you." There are two aspects of the seriousness question to be reported: extent of *inconvenient* acts and extent of *moral seriousness*.

Inconvenient acts are those which are disturbing to the milieu but which are without direct harm to either the actor or to others and do not violate an important moral taboo. Such acts seem to be natural expressions of children's motor and verbal needs. The flavor of this type of sin can best be conveyed by the following examples:

> School: "Talk and play." "Run around and make noise." "Throw paper." "Run in the halls."
>
> Home: "Play around at bedtime." "Talk at the table."

In the first grade inconvenient misconducts for school were reported by 42 per cent of the children, home 21 per cent ($p < .001$). In the third grade, inconvenient misconducts were reduced in both settings but the same preponderance of inconvenient sins for school occurred; school 21 per cent, home 11 per cent ($p < .01$).

While inconvenient categorization described a misconduct of a relatively less serious nature, it did not distinguish among the moderately serious and more severely serious misconducts. A scale of general moral seriousness was devised to extend over the whole range of moral seriousness. Results here showed that children in the first grade gave many more truly serious misconducts for home than for school ($p < .01$); by the third grade, however, these serious sins were reported with similar frequency for both home and school.

Influence of Milieu in Misconduct Explanations

A child may explain that a misconduct is bad because (a) it results in some harm to the sinner, (b) it harms other persons, or (c) it is just wrong. (Cases in which the reason given for misconduct being bad was that it was wrong were labeled *reflexive justifications.*) Results from this simple categorization of misconduct explanations are given in Table 2. (Since there were no statistical differences between boys and girls, the data for the sexes were combined.)

Inspection of Table 2 shows that concern for sinner as opposed to concern for others shows no influence of milieu. The only consistent distinction between home and school was that relating to reflexive justification; expla-

TABLE 2

CHILDREN'S EXPLANATIONS FOR WHY A SIN IS BAD

	Harm to Sinner	*Harm to Other*	*Reflex. Just.*
First Grade			
Home	120	81	26
School	110	73	44
Third Grade			
Home	95	113	19
School	82	113	32

nations of "it's bad because it's bad" were most frequently used for school misconducts ($p < .01$ and .05). Inspection of the data showed that the higher proportion of reflexive justifications at school was associated with their relatively greater use in connection with rule violations. In the first grade, for example, reflexive justifications were attached to 24 of the school rule violations, but to only four of the home violations. Third grade data showed the same trend.

It was possible to categorize reasons for misconduct badness more specifically. The frequently used explanations fell into the following categories:

Adults suffer. "My mother would have to clean up the mess."

Children suffer. "Kid could be hurt."

Adults punish. "I'd have to stay inside for a week."

Reality punishes—No person is made responsible for the punishment; the harm to the sinner is a natural outcome of his behavior. "If you don't look both ways you could get hurt real bad." "You could get behind in your work."

Reflexive Justifications and other.

(Sex differences were statistically insignificant, and the sexes were combined.)

Two milieu contrasts were consistent in Table 3. First, only home adults suffered as a result of a child's misconducts; teachers almost never suffered. Secondly, children at school suffered more frequently than children at home ($p < .001$). Apparently adults punished with similar frequency in both home and school. In the first grade, reality punishments are much more frequent for home than for school ($p < .001$). Reality punishments at home probably reflected the large number of dangerous situations at home about which younger children are constantly warned: playing in street, climbing tress, etc.

Just as milieu influenced whether or not the central adult suffered, it also influenced the type of punishment this central adult administered. In

TABLE 3

CHILDREN'S SPECIFIC EXPLANATIONS FOR WHY A SIN IS BAD

	Adults Suffer	*Children Suffer*	*Adults Punish*	*Reality Punishes*	*R. J. and others*
First Grade Children					
Home	42	37	50	58	40
School	10	59	54	26	78
Third Grade Children					
Home	60	46	28	64	29
School	1	92	16	54	64

the first grade, parents were reported as using corporal punishments in 64 per cent of their punishments, teachers 22 per cent ($p < .001$). By the third grade punishment by adults was markedly reduced, but home tended to maintain the greater proportion of corporal punishment: parents 32 per cent, teachers 19 per cent ($p < .10$).

Sex and Age Influences in Misconduct Type

For purposes of clarity, differences in misconduct ideas relating to sex or age have been so far ignored. However, certain findings indicate that both age and sex sometimes operated in relation to milieu influences and that age was occasionally relatively independent of milieu.

The conjunction of milieu, sex, and age variables is particularly striking in connection with the categories of misconduct in Table 1. In the first grade, boys' and girls' responses in Table 1 were quite similar for home ($p > .20$). First grade school, however, created sharp differences between boys and girls ($p < .01$). Here girls emphasized the school rules (often talking), and boys were concerned with assaults and with breaking. In the first grade, girls accepted the emphasis the school puts on rule violations; at this time and place the more aggressive concerns of assaults and breakings were diminished. However, by the third grade, the pattern of school misconduct for boys and girls was roughly similar ($p > .40$).

A second boy-and-girl difference relating to milieu was the girls' heightened concern with nonconformance to parents. In the first grade, girls had 26 home nonconformances to boys' 11; in the third grade, girls had 15 to the boys' 3 ($p < .01$ in both cases).

Boy-and-girl differences in *fire play* were impressive only for the home; although boys showed a weak tendency to have more of these at school than girls, it was only in the home responses that boys' *fire play* shows a decidedly greater frequency (p's $< .01$). Finally, as noted before, talking in school tended to be a girls' problem ($p < .01$).

The only impressive change in misconduct type related to age has to do with aspects of seriousness. As might be expected, inconvenient misconducts decreased with increasing age ($p < .01$), and general moral seriousness increased with age ($p < .01$); these results held for both home and school.

Sex and Age Influences in Misconduct Explanations

No consistent differences were found between boys' and girls' explanations as to why a misconduct was bad.

Differences in age did relate to misconduct explanations. It is clear from Table 2 that as the child gets older he increases his concern with harm to others. This change is consistent over settings: concern for others at home rises from 32 to 50 per cent, at school from 36 to 50 per cent (p's $< .01$).

If Table 3 is used to compare age changes, it can be shown that the increased concern about others at home resulted from an increase in adult suffering ($p < .05$), and the increased concern for others at school derived from an increase in child suffering ($p < .001$). Table 3 also shows that adult punishment as a reason for avoiding a misconduct is markedly reduced as children mature. Again the difference is consistent across settings: adult punishment at home drops from 22 to 12 per cent; at school, from 24 to 7 per cent (p's $< .01$). The type of punishment expected from adults also changes with age: corporal punishment for home drops from 32 to 9 per cent; for school, from 12 to 3 per cent ($p < .001$ and $.02$).

Since the data here presented are somewhat complicated, a listing is offered to summarize the major relations of milieu, sex, and age variables to the indices employed.

Milieu Contrasts:

School misconducts emphasized assaults on children and inconvenient and talking misbehaviors.

Home misconducts were weighted toward breaking of objects, fire play, and nonconformance with parents.

School explanations emphasized reflexive justification, children suffering, and noncorporal punishment.

Home explanations emphasized adult suffering and corporal punishment.

Sex Contrasts:

Girls emphasized talking at school; boys, fire play at home.

Girls also emphasized noncomformance with parents.

Girls in the first grade were dominated by concern with school rule violations while boys were more likely to give assaults and breakings; this was a temporary reaction which disappeared by the third grade.

Boys and girls did not differ in the type of explanation they offered for misconducts.

Age Contrasts:

As children became older, they increased their offering of morally serious misconducts and decreased stating inconvenient ones.

With age development, children's explanations referred more frequently to harm to others and less frequently to adult punishment; adult punishment, when expected, contained less corporal punishment.

DISCUSSION

Children probably answer the question, "What's the worst thing a child can do?" with a report of their preoccupations. It is not likely that our subject's answers would have been the same if they were presented with alternatives of different moral seriousness. Given a choice, most children would probably say that "stabbing someone" was worse than "talking while the teacher is out of the room."

If the misconduct ideas are taken to represent preoccupations, it should be possible to infer from these preoccupations the impact of the school or home milieu upon our subjects. Their preoccupations should reflect (a) the needs which they want to express but which the milieus find alien, (b) the temptations or opportunities which the milieu offers, and (c) the expectations held about what gets one into trouble. Preoccupations should change over time as some misconduct issues become settled and other issues take their place.

Using such an inferential approach with regard to milieu contrasts, it might be assumed that school is perceived as a place in which one's natural inclinations toward such things as motion and noise (inconvenient misconducts) are most heavily suppressed. The problem of handling one's aggressive impulses toward other children is also seen as particularly acute at school. There is probably a double reason for this: first, many other children are ever present at school, presenting a constant temptation—or opportunity; second, school will not tolerate the aggressive interactions, many of which are either excused or unnoticed at home.

Home is perceived as a place where a freer range of behavior is possible and where certain adventurous sins can result. At home a child has more occasion to be without close adult supervision; he can avail himself of more adult props and use them more freely. From this relative freedom come more misconducts dealing with fire play and breaking. One may also infer that parents and teachers occupy quite different roles in the child's mind. While teachers seem to be relatively unassailable (few nonconformances), parents seemed more vulnerable (more nonconformances).

Milieu influences also extend to explanations about why a misconduct is bad. There are indications that children have difficulty in seeing any good reason for some of the school rules. This deduction was supported by the fact that children reported more reflexive justifications for school, and these reflexive justifications were especially associated with school rule violations. Apparently, at school, one simply takes some prohibitions on faith.

The relatively unassailable role of teachers seems also related to misconduct reasons. Children rarely perceived school adults as suffering harm from their misconducts; they perceived home adults suffering such harm very frequently. Although the parent is seen as more vulnerable, the data on corporal punishment would indicate that he or she was also seen as inflicting more primitive, and probably more painful, punishment. One is not inclined to psychologize over this issue: parents most probably use corporal punishment more frequently than teachers.

One might hope that the school, as a socializing agent, would develop an increasing concern for others; this would be reflected in the present data by school explanations which showed more interest in harm to others and less concern about adult punishment. As reported, however, these improvements in socialization were a function of age and operated for both settings. What the school did do was to determine the kind of others with which the child became concerned: school increased children's concern with what happens to other children while the home increased concern with what happens to other adults. One might argue that the very large increase in child suffering at school represents a socialization gain superior to the rise in adult suffering at home. The reasoning here would be that a broadened concern about others of equal power (children) represents a more justice oriented outlook than increased interest in others of superior power (parents).

The results relating to sex differences indicated that these were few, specific, and related to misconduct type—not to misconduct explanations. One might have expected some impressive sex differences because it is commonly accepted that boys have more trouble adjusting to school than girls. In one large city, referrals to visiting teachers of boys with school conflict problems outnumbered referrals of girls by a ratio of five to two (1). One might expect school conflict among boys to result in boys offering a pattern of misconduct concerns different from the one offered by girls. This expectation was born out in the first grade school data: girls seemed to take on the school emphasis of obeying rules; boys became more concerned with assaults. However, by third grade, boys and girls showed no statistically significant difference in misconduct categories; girls reduced their concern with violations of rules and increased their assaults—they become like the boys. Whatever marks the difference between boys' and girls' adjusment in school, it seems unrelated to ideas about misconduct in school. Future research might investigate the question of whether this same sex similarity continues in the higher grades.

The constant sex differences of girls' concern with nonconformance at home and with talking at school and boys' interest in fire play at home are reported as matters of possible interest without attempts at explanation.

Age changes would indicate that, over time, children show their social maturation in two ways. First, they become able to see beyond the inconvenient sins that often get them into trouble to misconducts of more moral significance. Secondly, they begin to drop concern with obvious self-centered consequences, like adult punishment, and become more interested in what happens to other persons.

SUMMARY

Children in the first and, later, in the third grades were asked which misconducts they considered the worst at school and at home; they were also asked to explain why they considered these misconducts bad. Data were analyzed in terms of how relevant milieu was related to the qualities of the misconducts and to their explanations. Comparisons were made across variables of sex and age. Results showed that relevant milieu related both to the kinds of misconducts reported and to explanations for misconducts. Some of these relations were consistent regardless of the age or sex of the child; others obtained only for specific sex or age groups.

Differences in misconduct ideas related to sex were confined to kinds of misconduct—not to explanations—and were limited to those occurring in

one milieu or the other. Differences related to age tended to hold for both sexes and in both milieus.

Hypotheses were offered concerning the ecological and maturational variables responsible for the obtained relations.

REFERENCES

1. BAKER, H. J. *Visiting teacher program.* Psychological Clinic, Detroit Public Schools, 1958-59.
2. SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* McGraw-Hill, 1956.

MEASUREMENT OF CHILDREN'S SENSITIVITY TO AUDIENCES[1,2]

ALLAN PAIVIO,* ALFRED L. BALDWIN,
and SEYMOUR M. BERGER[3]

Cornell University

In our culture at least, social rewards and punishments are generally contingent on public exposure. As Heider (4) points out, what is perceived is often judged favorably or unfavorably. It is not surprising, therefore, that a public performance may involve elements of either attraction or fear, or both, for the "performer." Thus, we observe that some individuals "seek the limelight," others react with shyness, embarrassment or "stage fright" in the presence of observers and characteristically avoid public situations. Others apparently manifest both reaction tendencies to a high degree, e.g., professional actors who report extreme stage fright despite years of stage experience. Hollingworth (5) emphasized this conflictful nature of public exposure, referring to the "striking paradox [of] the joint activity of two opposing impulses—craving for an audience and fear of it" (5, p. 183). The "paradox" is significant if not surprising. Arousal of one or both of the opposing tendencies may be reflected in the quality of an individual's public performance as well as in his willingness to appear at all before an audience, and adequate measurement of the strength of these tendencies should enable us to predict individual differences in audience effects.

The present report describes the development and initial validation of a questionnaire designed to measure these conflicting tendencies in children. This questionnaire was initially based on an adult inventory developed by Paivio and Lambert (10) for the measurement of audience sensitivity, a concept defined in their study as a predisposition to be anxious in audience situations. However, in addition to describing anxiety and avoidance of public exposure, the children's questionnaire attempts to measure the motive to seek an audience, thus introducing the possibility of an approach-avoidance conflict analysis of behavior in audience situations. Such an analysis was employed in a separately-reported study (8), where children classified

---

* Department of Psychology and Sociology, University of New Brunswick, Fredericton, N. B., Canada.

[1] This research was supported by a grant (M-901) from the National Institute of Mental Health.

[2] We wish to thank Dr. Mary Gallwey and Irene Stein for their help with aspects of the study.

[3] Now at Indiana University.

Child Develpm., 1961, 32, 721-730.

(on the basis of scores on the children's questionnaire) as having strong tendencies both to seek and avoid public exposure made more speech "errors" before an audience than did less conflicted subjects. Further theoretical considerations and related research are more fully discussed elsewhere (7, 9, 10).

DEVELOPMENT OF THE QUESTIONNAIRE

Initial Form of the CASI

The adult Audience Sensitivity Inventory (10) consists of 21 items referring to typical reactions to various audience situations. Items from this scale were reworded to be more suitable for children and some new items added, forming a 31-item Children's Audience Sensitivity Inventory (CASI).

The CASI was mimeographed and group-administered to 111 fourth-, fifth-, and sixth-grade children from an urban elementary school. Procedural instructions and the items were read aloud by an *E*, and the children responded by checking one of two answer spaces (labeled "Yes—like me" and "No—not like me") following each item.

Twenty of the items showed enough variation to be subjected to further analysis. A matrix of phi coefficients, computed between the dichotomously-scored items, was factor analyzed to simple structure, following Thurstone's (13) centroid method. The analysis yielded two clearly orthogonal factors. The first showed high loadings on seven items referring to public exposure of the self or a self-product (e.g., I like to tell a story in front of the class, I like to show my work to my classmates, etc.) and was labeled Exhibitionism (Exh). The second was defined by six items referring to negative attitudes toward public exposure (e.g., I feel bad when someone sees me goof) and was called Self-Consciousness (S-C).

Assuming that the Exh factor measures an audience-seeking tendency and S-C apprehension with respect to public exposure, orthogonality of the factors meant that *S*s scoring high on both (i.e., *S*s for whom a public performance presumably involves an approach-avoidance conflict) could readily be selected for research purposes. However, examination of the two factors suggested caution in their interpretation. The items comprising Exh are relatively specific in their referents, whereas the S-C items are more global and less situation-specific, suggesting that the factors may not be of comparable validity as predictors of reactions to particular audience situations. Some evidence that this might be the case was obtained in a separate study (9) concerned with the antecedents of audience sensitivity. One hundred and eighty-seven third and fourth grade children wrote compositions on either "Why I like to recite in front of the class" or "Why I do not like to recite in front of the class," the children being free to choose either theme. Liking reciting, as expressed by the composition theme, correlated significantly (biserial $r = .67$, $p < .0001$) with Exh, but only negligibly with S-C. In itself this finding suggested only that the composition theme could be interpreted as a further Exhibitionism "item." However, analysis of the content of the compositions revealed that 85 per cent of the compositions written on "Why I do not like to recite" included statements clearly referring to anxiety as the reason for not liking reciting (e.g., "because I shake all over," "because I am scared," etc.). Thus, while Exh correlated negatively with direct statements of anxiety concerning recitation, S-C was apparently unrelated to such references and could not be unequivocally interpreted as measuring fear-mediated avoidance of public exposure.

The above considerations led to a revision of the CASI to include a third scale designed to measure situationally-specific avoidance tendencies regarding public performance.

Revised CASI

Items. Sixteen items which refer explicitly to anxiety reactions in, or avoidance of, the kinds of situations represented by the Exh items were added to the CASI. Some of these were direct modifications of Exh items to include a fear reference; others were reproductions of popular statements made by *Ss* who wrote compositions on "Why I do not like to recite" in the study (9) cited above. The new "scale" was labeled Audience Anxiety (AA).

In the course of the revision, six new items were added to the Exh factor to increase its reliability. The revised CASI thus consists of 13 Exhibitionism, 6 Self-Consciousness, and 16 Audience Anxiety items. The items are listed below.

Exhibitionism items:

I like to recite poems in front of other people.
Everytime I get a chance to do something in front of the class I take it.
In school, I always raise my hand if I know the answer.
Even if I know the answer, I usually do not raise my hand. (—)[4]
If I wrote a prize-winning poem I would rather have someone else read it in front of the class. (—)
If my paper is hung on the bulletin board, I'd rather *not* have my name on it. (—)
I like to show things I make to other children.
I would like to make something while the whole class watches.
I like to show my work to my classmates.
I like to sing in front of others.
I like to tell a story in front of the class.
I would like to be on the stage in front of many people.
If my paper is hung on the bulletin board, I like everyone who sees it to know it's mine.

Self-Consciousness items:

I am bashful with most strangers.
I feel funny when I walk into a room full of people.
I often wonder what others think of me.
I feel bad when someone sees me goof.

---

[4] Items followed by a minus sign (—) are scored if the response is negative.

I often worry about what people think of me.
Other people can hurt my feelings easily.

Audience Anxiety items:

I worry that I might forget my lines when I recite poems in front of the class.
I do not like acting in plays because I am afraid that people will laugh at me.
I feel nervous when others look at work I have done.
I feel embarrassed when I am asked to sing in front of others.
My knees shake when I recite in class.
I worry that I will make a mistake when I give a report in front of the class.
I never feel one bit nervous when I recite. (—)
When it is my turn to get up and recite in class, I feel my heart pounding hard.
I do not like reciting in class because I might make a mistake and others would laugh at me.
I am afraid that other children will laugh at me when I show my work to them.
Sometimes I shake all over when I am asked to recite in class.
Sometimes my voice shakes when I recite in class.
Once I had to do something in front of an audience and I was really scared.
I feel scared when I recite in front of the class.
I am very nervous if the whole class watches me while I am making something.
When I recite in class I often wonder what others are thinking about me.

Reliability of the revised CASI. Test-retest (six-month interval) and split-half reliabilities of the Exh, S-C and AA scales were computed for a sample of third-, fourth-, and fifth-grade children from two elementary schools. The questionnaires were group-administered by teachers, following detailed instructions provided by an *E*. The reliabilities are presented in Table 1.

TABLE 1

TEST-RETEST AND SPLIT-HALF RELIABILITIES OF THREE SCALES (THIRD-, FOURTH-, AND FIFTH-GRADE CHILDREN)

	N	*Exh*	*S-C*	*AA*
Test-retest *r*[*]	209	.62	.45	.54
Split-half *r*[†]	223	.67	.66	.80

[*] Six-month interval.
[†] Corrected by Spearman-Brown formula.

The test-retest correlations are low but not necessarily unsatisfactory, considering the long retest interval. They may in part reflect intervening classroom experiences which affected the motivational predispositions the scales were intended to measure. The problem requires separate investigation, however, and extensive speculation is unnecessary here.

Correlations between CASI, Manifest Anxiety, and Test Anxiety Scales. While mainly concerned with the relations among the three CASI scales, we were also interested in comparisons with instruments on which published data were available. Thus, the items of the children's forms of the Manifest Anxiety (CMAS) (1) and Test Anxiety (TA) (11) Scales were added to the CASI. The composite questionnaire was administered to three samples of children from school grades 3 through 10 (as part of separate studies, the purposes of which are not relevant here), including the sample on which the above reliability data are based. Intercorrelations of the five scales were computed, separately for boys and girls, for each of the samples. In some instances there is considerable variability in the individual sample correlations between pairs of scales, but the over-all pattern of relations among the five scales is similar in the different samples and their data will not be individually presented here. In order to give a representative picture of the interrelations, the data were pooled (total $N = 421$) using Fisher's z'-transformation method (*see* 3) for computing average correlations. These correlations are presented in Table 2.

TABLE 2

AVERAGE INTERCORRELATIONS AMONG FIVE QUESTIONNAIRE SCALES

	Exh	*S-C*	*AA*	*TA*	*CMAS*
Exh		—.21	—.55	—.40	—.27
S-C			.58	.59	.61
AA				.74	.62
TA					.70

NOTE.—Averages are based on six subgroup *r*s (three samples, boys and girls treated separately) and computed by Fisher's z'-transformation method (total $N = 421$). Homogeneity of subgroup *r*s not demonstrated in every case. All average *r*s are significant, $p < .01$.

Considering, first, the relations among the CASI scales, Table 2 shows that Exh and S-C are only slightly negatively correlated, as in the factor analysis of the initial CASI. However, audience anxiety is apparently a component of both Exh and S-C, for AA correlates moderately with each (i.e., negatively with Exh, positively with S-C). To the extent that such inferences are warranted from questionnaire data, the Exh-AA correlation suggests that attraction and fear vis-à-vis public performance are neither bipolar nor independent. A subject could score moderately highly on both scales, presumably indicating that he is somewhat conflicted in his attitude toward audiences, but he is unlikely to be at once extremely exhibitionistic and highly fearful of public exposure. Thus, considering at this point only the intercorrelations of the CASI scales, an approach-avoidance conflict analysis of audience-oriented behavior would be possible using either AA or S-C scores as measures of a dispositional avoidance tendency, but would clearly be more feasible with the latter in view of its relative independence of Exh.

The correlations among the anxiety scales and S-C are all moderately high. Most notable among these (particularly since the two scales were independently developed for somewhat different purposes) is the correlation between AA and TA, which consistently ranges between .70 and .80 for the different subsamples (average $r = .74$). This and other aspects of the questionnaire data will be further discussed at a later point in the paper.

BEHAVIORAL VALIDATION OF THE REVISED CASI

The following study, undertaken to test the validity of the CASI scales, takes advantage of a natural setting in which children were asked to volunteer for a public performance. It was assumed that, if the Exh factor is a valid measure of an audience-seeking tendency, high scorers on Exh should volunteer more readily for a public performance than *S*s scoring low on Exh. Similarly, to the degree that the S-C and AA scales measure the tendency to avoid public exposure, scores on these scales will correlate negatively with "willingness to volunteer."

Subjects

The *S*s were 56 boys and 54 girls attending (at different times) a summer camp near Ithaca, N.Y. The boys ranged from 10 to 16, the girls from 8 to 15, years of age.

Procedure

Boys. The boys' group consisted of five camp units, the number of boys in a unit ranging from 8 to 18. Their camp program included a "skit" night for which each unit was to prepare a public performance and for which the campers were asked by their unit leaders to volunteer. Participation was optional, but considerable social pressure was presumably involved. The unit leaders were asked to rate each boy in their respective units on eagerness to volunteer for the skit, using scoring sheets containing the following instructions:

> Please list below the names of those boys who volunteered to perform in a skit, and indicate, to the best of your knowledge, how *readily* each boy volunteered by writing the appropriate number (1, 2, 3, or 4) in the column headed "Willingness to Volunteer," where
>
> 1 = volunteered readily (appeared eager to participate)
> 2 = volunteered (but not particularly eagerly)
> 3 = volunteered reluctantly (needed some coaxing)
> 4 = volunteered *very* reluctantly (group pressure required)

Not volunteering at all represented a fifth category.

The CASI, CMAS, and TA scales were completed by each boy on the day of the skits. They were administered by the camp counselors who followed mimeographed instructions concerning the procedure.

Girls. A procedure similar to the boys' was followed with the girls' sample (four camp units of 10 to 18 *Ss*), but more carefully controlled. The skit night was included in their camp program at the request of an *E* who was able to meet with the girls' counselors and discuss the study in detail, and thereby establish uniformity in the procedure. Criteria were suggested for rating the children: *Ss* who volunteered immediately after the first call for volunteers were to be scored 1 ("Volunteered Readily"), those who agreed to participate after a second and a third request were scored 2 and 3 respectively, and those who volunteered only after the greatest insistence by the counselor and other campers were scored 4 ("Volunteered very reluctantly").

The questionnaires were completed by the girls on the day following the skits.

Results

It would have been desirable to consider each camp unit individually, but this was impractical with the boys' data. In view of the small numbers in some of their groups and the fact that two of the counselors used only a two-point scale in rating the boys on volunteering (1 and 2 in one case, 1 and 5 in the other), the boys' volunteering scores were dichotomized at a point where a common division was possible and the units combined. The resulting two groups to be compared consisted of 27 *Ss* rated as "volunteering readily" and 29 rated as volunteering less readily or not at all. Biserial correlations were computed between these dichotomized volunteering scores and each of the CASI scales.

While the procedure for the girls' study differed somewhat from that of the boys, the conditions for the two studies seemed sufficiently similar to make comparisons meaningful. Thus, the girls' data were analyzed in the same manner as those of the boys. The resulting biserial correlations between willingness to volunteer for the skits and CASI, TA, and CMAS scores, for both sexes, are shown in Table 3.[5]

All correlations are in the expected direction, i.e., volunteering for public performance correlates positively with Exh and negatively with S-C and AA (as well as with CMAS and TA).[6] However, conventional significance

---

[5] Since all but one of the girls' counselors used the four categories of the scale in rating the *Ss* in their units on volunteering, an alternative analysis was feasible with the girls' data. Product-moment correlations between the questionnaire scores and volunteering ratings were computed for each of the groups and combined into single correlation estimates, after establishing homogeneity of the individual *r*s, following Fisher's (3) technique. The resulting average correlations are generally comparable to the girls' biserial correlations shown in Table 3 and will not be separately discussed.

[6] In view of their intercorrelations (Table 2) the scales are obviously not independent predictors of volunteering.

TABLE 3

BISERIAL CORRELATIONS BETWEEN RATED "WILLINGNESS TO VOLUNTEER FOR PUBLIC PERFORMANCE" AND SCORES ON FIVE QUESTIONNAIRE SCALES

Sex	*N*	*Exh*	*S-C*	*AA*	*TA*	*CMAS*
Boys	56	.31*	—.23	—.35*	—.36*	—.29*
Girls	54	.13	—.20	—.45**	—.39**	—.26

* $p < .05$, one-tailed test.
** $p < .01$, one-tailed test.

levels are not reached in every case. The AA and TA correlations with volunteering are significant ($p < .025$, one-tailed test) for both boys and girls, Exh and CMAS for boys only ($p < .05$), while the S-C correlation is not significant in either sample. It should be noted, however, that none of the sex differences is significant, increasing confidence that correlations in the predicted direction obtain.

Particularly noteworthy is the correlation of $-.45$ ($p < .01$) between girls' AA scores and volunteering, which indicates that at least 20 per cent of the variance in rated volunteering scores is accounted for by "audience anxiety."

DISCUSSION

The results of the preliminary validating study generally support the validity of the Children's Audience Sensitivity Inventory. The accuracy of prediction leaves much to be desired from a practical point of view, but the Audience Anxiety scale correlations with volunteering (particularly in the girls' sample) compare favorably with validity coefficients of personality tests in general (*see*, e.g., 6). The Exhibitionism scale was less successful as a predictor, only the boys' correlations with volunteering being significant. Possibly the Exh items are less reliable predictors for girls because skill in the activities they imply (singing, reciting, acting, etc.) is more valued by girls (cf. 12) and, hence, social acceptability plays a more important role in determining girls' than boys' responses to the items. There may also have been some effect from the fact that the CASI was completed by the boys before, and by the girls after, actual participation in the skit; i.e., experience in the skit may have affected the girls' responses to the Exhibitionism items but could not have done so in the case of boys. However, as already pointed out, the sex differences in the correlation between Exhibitionism and volunteering is not significant, and extensive speculation is unnecessary.

The Self-Consciousness scale also did not correlate as highly with volunteering as did Audience Anxiety. The difference may be due to the smaller number of items in the S-C scale and/or to the specific relevance of the AA items to the criterion situation in the present study. In any case, it does not justify assuming lower validity of the S-C items with respect to other kinds of audience situations.

The relation between Audience Anxiety and Test Anxiety is of some interest here. The high correlation between them and the fact that, of the scales included in the questionnaire, they correlate most highly (and approximately equally) with volunteering, suggests common anxiety-arousing features in the situations implied by the two sets of items. The crucial element may simply be the classroom situation, to which most items of both scales make reference, but the fact that the criterion situation in the present study was *not* a classroom suggests something beyond this.

Both the test and the audience situation involve, implicitly or explicitly, evaluation by others. Test Anxiety and Audience Anxiety scores may therefore reflect a generalized "sensitivity" to social evaluation founded on prior experience in which behavior has been devalued and punished. Such an interpretation is supported by the findings of a study on antecedents of audience sensitivity (9): children whose social behavior and achievements were evaluated relatively unfavorably by parents, and who were punished frequently for failure to meet parental standards, had significantly higher AA (and TA) scores than children who were favorably evaluated and infrequently punished. The same interpretation may hold for the Manifest Anxiety Scale, which contains a number of items having reference to social evaluation (e.g., "I feel that others do not like the way I do things"). However, while moderately intercorrelated, indicating some general factor, an item analysis (not reported here) shows that the items of each scale correlate more highly with their "own" scale than with the others, suggesting situational specificity in anxiety (cf. a study by Dixon *et al.* [2], in which a factor analysis of a "social anxiety" scale yielded a strong general factor and four group factors). The basis of such specificity cannot be determined from existing data.

The concepts employed in this study raise many questions, both regarding their interpretation and their relation to such concepts as achievement need, self-esteem, etc. Such questions and specific lines research might take to answer them are discussed elsewhere at length (7, 9) and shall not be introduced here. Ultimately, the value of these concepts depends on how far they will take us in the description and conceptualization of social influence, particularly that aspect of social influence which is attributable to observing, evaluating, and potentially approving or disapproving others. The instruments described above have already proven useful in that regard (*see* 8, 9), and further research is currently under way.

SUMMARY

A Children's Audience Sensitivity Inventory was developed to measure individual differences in motivational reactions to the presence of an audience. The Inventory includes three scales labeled Exhibitionism, Self-Consciousness, and Audience Anxiety.

Empirical validation of the inventory involved a camp situation in which children volunteered for performance in a skit. It was predicted that Exhibitionism would correlate positively, and Self-Consciousness and Audience Anxiety negatively, with willingness to volunteer (as rated by camp counselors). The findings generally supported the predictions, most strongly in the case of Audience Anxiety. Relations to the children's Test Anxiety and Manifest Anxiety scores are also discussed.

REFERENCES

1. CASTANEDA, A., MCCANDLESS, B. R., & PALERMO, D. S. The children's form of the manifest anxiety scale. *Child Develpm.*, 1956, 27, 317-326.
2. DIXON, J. J., DE MONCHAUX, C., & SANDLER, J. Patterns of anxiety: an analysis of social anxieties. *Brit. J. med. Psychol.*, 1957, 30, 107-112.
3. EDWARDS, A. L. *Experimental design in psychological research.* Rinehart, 1956.
4. HEIDER, F. *The psychology of interpersonal relations.* Wiley, 1958.
5. HOLLINGWORTH, H. L. *The psychology of the audience.* American Book, 1935.
6. KIMBLE, G. *Principles of general psychology.* Ronald, 1956.
7. LEVIN, H., & BALDWIN, A. L. Pride and shame in children. In M. R. Jones (Ed.), *Nebraska symposium on motivation.* Univer. of Nebraska Press, 1959. Pp. 138-173.
8. LEVIN, H., BALDWIN, A. L., GALLWEY, M., & PAIVIO, A. Audience stress, personality and speech. *J. abnorm. soc. Psychol.*, 1960, 61, 469-473.
9. PAIVIO, A. Child rearing antecendents of audience sensitivity. Unpublished doctoral dissertation, McGill Univer., 1959.
10. PAIVIO, A., & LAMBERT, W. E. Measures and correlates of audience anxiety ("stage fright"). *J. Pers.*, 1959, 27, 1-17.
11. SARASON, S. B., DAVIDSON, K., LIGHTHALL, F., & WAITE, R. A test anxiety scale for children. *Child Develpm.*, 1958, 29, 105-113.
12. TERMAN, L. M. Psychological sex differences. In L. Carmichael (Ed.), *Manual of child psychology.* Wiley, 1946. Pp. 954-993.
13. THURSTONE, L. L. *Multiple factor analysis.* Univer. of Chicago Press, 1947.

PERSONALITY AND BEHAVIOR PATTERNS OF CHILDREN WITH FUNCTIONAL DEFECTS OF ARTICULATION[1]

ARTHUR L. SOLOMON*

Chico State College

It is reported that the outstanding majority of speech defects in children are articulatory in symptom and functional in origin (11, 17), although no clear-cut etiological picture has as yet been established to account satisfactorily for the occurrence of the defects or for their subsequent disappearance. Studies of somatic variables over the past 20 years have not been especially rewarding (13), while clinical evidence has increasingly stressed the emotional and personality factors found as concomitant symptomatology in children with various functional speech disorders described as "baby talk," "infantile perseveration," "delayed speech," "speech inhibition," or "mutism" (1, 3, 4, 6, 12, 15).

Experimental studies using personality inventories have been either inconclusive or contradictory, although the investigations of Wood (16), Nelson (10), and FitzSimons (5) of children with functional defects of articulation, using rating scales, controlled observations, or projective techniques, have yielded more conclusive findings. Beasely (2), McCarthy (7), and Mowrer (9), particularly, have emphasized the importance of the parent-child rela-

---

* Speech and Hearing Clinic, Chico State College, Chico, California.

[1] Adapted from a doctoral dissertation, Stanford University, 1960. The author wishes to acknowledge the generous assistance of Dr. Virgil A. Anderson, Director of the Division of Speech Pathology and Audiology of the Stanford School of Medicine; Dr. Dorothy Huntington; and Dr. Lois Meek Stolz, Professor of Psychology Emeritus, Stanford University.

Child Develpm., 1961, 32, 731-737.

tionship in the acquisition and adequacy of language, as well as other social learnings.

It would appear to follow that disturbances in the socialization process due to certain unfavorable environmental factors and showing up as problem behavior or unfavorable personality traits would, because of the sensitivity of the speech mechanism to disturbances in interpersonal relationships, be reflected in communicative disturbances as well. This hypothesis is apparently supported by the observation that speech defects of a functional nature follow patterns very similar to those of the usual childhood behavior problems which reach their ascendancy in the preschool years and gradually diminish or become displaced during the early years of school (8, 14). If these implications are correct, behavior problems, or specific personality traits and their relation with speech disturbance, should then be able to be experimentally demonstrated. If demonstrated, some further evidence will be added to the supposition that functional speech defects in children are part of a total adjustive pattern and not disparate from other social learnings.

A study was consequently designed to explore the behavior and personality traits of children with functional defects of articulation to determine whether children with these defects differed from their normal-speaking peers in specific behavior problems; whether the severity and degree of that difference could be determined; whether the behavior and personality traits of these children followed any particular pattern; and whether there might exist a correlation between the severity of the articulatory defects and the severity of the behavior disturbances, if any.

METHOD

Forty-nine boys and girls, of a comparatively homogeneous socioeconomic background, with functional defects of articulation, were matched in pairs with a control group of normal-speaking children on the basis of age, grade, sex, and intelligence. The experimental group represented a broad sampling of the speech defective population as no selection was made on the basis of severity of speech involvement. For inclusion in the experimental group, the subjects had to exhibit at least one consistently defective sound; they had to have scored 90 or above on the nonverbal scale of the California Short-Form Test of Mental Maturity: they had to be in the first grade; not a member of a twin pair; they had to have good health, good hearing, and a negative medical history; they had to have normal speech structures; and a mother who was willing and able to participate in the study. For inclusion in the control group, the subjects had to have normal articulatory proficiency and had to be within 90 calendar days and within five points of their experimental counterparts in age and IQ, respectively, and of the same sex.

In order to ascertain the existence and degree of problem behavior and to evaluate personality differences, mothers of both groups were interviewed and asked to describe their children's behavior, over the past two years, in nine representative areas of child development and personality. These areas consisted of (a) eating behavior, (b) sleeping behavior, (c) toilet training, (d) fears and anxieties, (e) comfort patterns, (f) tension, (g) aggression, (h) dependency, and (i) peer relations. A parent interview-questionnaire was devised to elicit information in these areas and, in general, the interviews followed similar patterns: (a) a general question to encourage the parent to describe her child's behavior in a given category and (b) specific questioning to determine frequency or severity of the reported behavior.

The interviews were recorded on tape and later typed verbatim from the recordings and so arranged that each of the nine behavior areas could be detached and rated separately. The protocols were rendered anonymous, and experimental and control data were randomized so that two judges, working independently, could rate all the children on each specific behavior area and again on each child's total behavior to determine an "over-all adjustment score." The ratings of the two judges, using a five-point scale, were averaged, yielding a combined score. Coefficients of agreement between the ratings of the judges were calculated along with a test of the consistency and range of their respective ratings. The investigator was one of the two judges participating in each rating procedure.

To determine the relationship between articulation and adjustment, the experimental subjects were divided by the investigator into two categories on the basis of their articulatory errors: (a) those with mild-moderate speech involvement and (b) those with marked-severe involvement. Then, these two groups were compared on their over-all adjustment scores. To determine characteristic behavior patterns, four categories were decided upon that might yield significant data, namely: (a) orality, (b) passivity-submissiveness, (c) internalizing behavior, and (d) externalizing behavior. The items under "orality" included thumb or finger sucking, nail biting, excessive crying, chewing objects, etc. The second category of "passivity-submissiveness" contained such examples of behavior as the following: tries hard to please, to be good, or to gain approval; exhibits excessive shyness and reserve; gives up easily; wants to be dressed, washed, fed, or fondled; seldom exhibits anger; and, is a follower. The third category of "internalizing behavior" included such behavior as restlessness in sleep, enuresis, embarrassment about toilet habits, nightmares, fearfulness, oversensitivity, perfectionistic tendencies, overconcern with cleanliness, nervousness, worries, and insufficient appetite. The fourth category of "externalizing behavior" consisted of destructiveness, displays of temper, defiance, cruelty, extreme domineering tendencies, and overactivity.

The critical ratio technique was used throughout to determine the significance of mean differences between the ratings for the control and experimental subjects on (a) specific behavior scores, (b) over-all adjustment scores, (c) relation between the severity of speech defect and the severity of problem behavior, and (d) characteristic behavior patterns that might differentiate the two groups. In those cases where correlations were computed, the Pearson product-moment method was used. The correlations of the ratings of the two judges ranged from .70 to .89 with a median correlation of .78 in the nine behavior areas and .67 for the over-all adjustment rating. Neither judge had any consistent tendency to rate higher or lower than the other.

RESULTS

Of the nine selected behavior areas studied, four areas, in addition to the over-all adjustment rating, were found to show differences between the experimental and control groups which were significant at the .01 level of confidence or better. From mothers' reports, the behavior areas in which the experimental group showed significant differences from their controls at the .01 level were sleeping, fears and anxieties, and tension; and, at the .001 level, in the area of peer relations and in over-all adjustment. No significant differences were found between the two groups of children in the areas of eating, toilet training, comfort patterns, or aggression.

In the area of sleeping behavior, the problems were centered around difficulties in falling asleep, restlessness, and other disturbances during sleep. In the area of fears and anxieties, the problems were centered around specific fears, areas of overconcern, unusual sensitivity, worries, withdrawal, and perfectionistic tendencies. In the area of tension, the children were considered by their mothers to be restless, sensitive, or nervous children who became upset and cried easily. They exhibited overt tensional outlets in such behavior as nailbiting, eyeblinking, pulling or twisting the hair, and in general overactivity. The experimental children were reported to have more disturbances in the area of peer relations. These problems were expressed in such behavior as a need to be first in games, a preference for children other than the child's own age or sex, an attempt to win other children's friendship by submissiveness or overgenerosity, a reluctance to stand up for one's self, and a preference for playing alone.

In addition to being rated as having poor "over-all adjustment" in comparison with the control children, certain personality traits were found to characterize the experimental children. The speech-defective group tended to be passive children who internalized their responses and were characterized by submissiveness, timidity, and a need for approval. Although there was a tendency for the severity of problem behavior to be related to the severity of speech involvement, the relation did not yield a significant level of confidence.

It is of interest to note that in no instance was the control group of normal-speaking children reported to have more problem behavior than the experimental group. They were reported as having similar problems, but to a lesser degree than the experimental children. However, it should also be noted that a sizeable number of children of both groups were found to have only slight problems, or none at all, in many of the areas under study.

DISCUSSION

The findings of this study tend to support the contention that articulatory speech problems of a functional nature are part of a larger picture of emotional and social behavior in a child's development. Based on reports by the mothers, the speech defects tended not to occur in isolation but frequently with other disturbances touching several areas of their social and personal adjustment, although this was by no means typical of every child in the experimental group.

The sleeping disturbance, the tension-reducing outlets, the fears, the inadequacies in peer relationships, plus the submissiveness and need for approval, suggest the possibility that more experimental children were operating under a greater degree of emotional stress than was true of their controls. A possible explanation for the existence of this stress may be hypothesized as being a generalized anxiety underlying the diverse symptomatology presented by many of the experimental children. If speech represents a form of self-assertion, then to "speak up" in the adult manner may be a frightening and forbidding prospect to apprehensive or submissive children who may have a need to display more infantile and nonassertive behavior in their particular environments. The speech defect in such cases may serve as a means of reducing anxiety and stress and reflect, as well, a fear of taking forward steps toward maturity and increasing independence.

Bearing in mind that similar problems were found with some frequency in both the experimental and control groups, it can be assumed that the dividing line between the behavior patterns and personality of children with speech problems of this type and children with normal speech remains a tenuous one, although statistically verifiable. The similarities between many of the children from both groups may be due to the fact that the experimental group contained children with varying degrees of speech involvement, some of which would be barely perceptible to the untrained ear. Another consideration that might add to the similarity of the two groups may be the fact that no child in our culture escapes entirely from some adjustive difficulty in his development toward emotional and social maturity. These factors may have tended to minimize behavioral differences between the two groups. Nevertheless, this study does suggest that more serious problem behavior of a particular kind, as well as certain specific personality characteristics, will be more apt to occur among children with functional defects of articulation than among a comparable group of normal-speaking children.

Because of the personality and adjustment problems of the speech handicapped child, rapport with the child, the development of self-confidence, and the creation of an atmosphere that encourages the child toward self-expression and self-assertion may be therapeutically more valuable than the specific speech correction measures employed. Prognosis in any given case will depend more explicitly upon the personality and the approach of the speech therapist and upon the quality of the relationship established with the child than has heretofore been generally recognized in practice.

SUMMARY

The purpose of the study was to test for a possible relationship between functional articulatory speech defects in children and personality and behavior patterns. Using an open-ended questionnaire type interview with mothers, the personality and behavioral background of a speech-defective experimental group and a normal-speaking control group was obtained in nine areas significant of child development and personality. The protocols were rated by two judges to obtain quantitative and qualitative differences in behavior, patterns of personality traits, and the correlation of speech and behavior problem severity.

From mothers' reports, the experimental group showed significant differences from their matched controls in their over-all adjustment, peer relations, sleeping behavior, manifestation of fears and anxieties, and tensional outlets. In personality, the experimental subjects tended to be passive children who internalized their responses and who were further characterized by submissiveness, timidity, and a need for approval.

Functional speech problems of this type, apparently, are not isolated phenomena but part of a total adjustive pattern. It is suggested that an underlying stress may be common to the diverse symptomatology here presented and that the infantile and nonassertive behavior could very well serve as anxiety-reducing devices to meet environmental pressure.

Therapeutically, rapport with the child, the development of self-confidence, and the creation of an atmosphere and relationship that encourages the child toward self-expression and self-assertion may be more valuable than the specific speech correction measures employed. In turn, these findings suggest a broadening of the speech therapist's training to include a firmer background in child psychology and development and in counseling and guidance techniques for both children and adults.

REFERENCES

1. ALLEN, I. M. Defect of the speech function in childhood. *New Zealand med. J.*, 1947, 46, 297-307.
2. BEASELY, J. Relationship of parental attitudes to development of speech problems. *J. Speech Hearing Disorders*, 1956, 21, 316-321.
3. BROWN, F. W. Baby talkers. *Proc. Amer. Speech Corr. Ass.*, 1936, 7, 197-208.

4. Chapin, A. B., & Corcoran, M. A. A program for the speech-inhibited child. *J. Speech Hearing Disorders,* 1947, 12, 373-376.
5. FitzSimons, R. Developmental, psychosocial, and educational factors in children with nonorganic articulation problems. *Child Develpm.,* 1958, 29, 481-489.
6. Irwin, R. B. Speech comes to a five year old boy. *J. Speech Hearing Disorders,* 1946, 11, 197-203.
7. McCarthy, D. Language disorders and parent-child relationships. *J. Speech Hearing Disorders,* 1954, 19, 514-523.
8. Macfarlane, J. W., Allen, L., & Honzik, M. P. *A developmental study of the behavior problems of normal children between twenty-one months and fourteen years.* Univer. of California Press, 1954.
9. Mowrer, O. H. *Psychotherapy, theory and research.* Ronald Press, 1953.
10. Nelson, O. W. An investigation of certain factors relating to the nature of children with functional defects of articulation. *J. educ. Res.,* 1953, 47, 211-216.
11. Powers, M. H. Functional disorders of articulation—symptomatology and etiology. In L. E. Travis (Ed.), *Handbook of speech pathology.* Appleton-Century Crofts, 1957. Pp. 707-768.
12. Russell, C. M. Personality factors in a motor speech delay case. *Amer. J. ment. Def.,* 1944, 49, 171-176.
13. Spriestersbach, D. C. Research in articulation disorders and personality. *J. Speech Hearing Disorders,* 1956, 21, 329-335.
14. Van Alstyne, D., & Hattwick, L. A. Behavior patterns of children with nursery school experience. In R. G. Kuhlen & G. G. Thompson (Eds.), *Psychological studies of human development.* Appleton-Century-Crofts, 1952. Pp. 411-415.
15. Werner, L. S. Treatment of a child with delayed speech. *J. Speech Hearing Disorders,* 1945, 10, 329-334.
16. Wood, K. S. Parental maladjustment and functional articulatory defects in children. *J. Speech Hearing Disorders,* 1946, 11, 255-275.
17. American Speech & Hearing Association Committee on Midcentury White House Conference. Speech disorders and speech correction. *J. Speech Hearing Disorders,* 1952, 17, 129-137.

EFFECTS OF PRETRAINING ON THE REINFORCING VALUE OF VISUAL STIMULI[1]

HAROLD W. STEVENSON *and* RICHARD D. ODOM*
University of Minnesota

Although there have been numerous studies indicating that visual stimuli may operate as effective reinforcers for the behavior of subhuman species and human adults, there is little information concerning this problem with children. The present study seeks to investigate the effect of visual reinforcement on children's performance in a simple operant task. In addition, the question is raised whether visual reinforcers are subject to satiation and deprivation effects.

The effects of two pretraining conditions on performance are investigated. *Ss* in one pretraining condition were left alone for 15 minutes in a room equipped with a large variety of novel and interesting mechanical toys. In a second pretraining condition, *Ss* spent the 15 minutes alone in the room with no toys. It is assumed that the effectiveness of visual reinforcers would be reduced following experience with the toys because of a partial satiation of a need for visual stimulation. It is also assumed that isolation with no toys would increase such needs and would thereby increase the effectiveness of visual reinforcement.

METHOD

Subjects

Ss were 30 boys and 30 girls attending kindergarten, first and second grades of the Laboratory School of the University of Minnesota.[2] All children in attendance on the days of the study were used as *Ss*. *Ss* were above average in intelligence and socioeconomic status. Six *Ss* were discarded for reasons described below and were replaced by other *Ss*.

---

* Institute of Child Development, University of Minnesota, Minneapolis 14.

[1] This study was supported by a grant from the National Science Foundation (Grant G 9687) to the first author.

[2] The writers wish to express their thanks to the staff of the University Laboratory School for their cooperation in this study.

Child Develpm., 1961, 32, 739-744.

Apparatus

Experimental task. The apparatus consisted of a wooden frame 20 in. long and 14 in. deep upon which two plywood panels were mounted. Centered on the front of the vertical panel was a rectangular aperture into which a 2 in. high and 3 in. wide piece of milk glass was inserted. The second panel sloped from a front elevation of 3½ in. to a rear elevation of 6½ in. Centered on the sloping panel was a vertical wooden lever, 2½ in. high and ¾ in wide. The apparatus was painted a flat gray with the exception of the horizontal panel and lever which were painted flat black.

Behind the apparatus was a 16 mm. Dunning Animatic filmstrip projector which projected a 1¾ in. by 2 in. picture on the milk glass screen. The apparatus and the projector were placed on a small table in a 5½ ft. square room. A dark green screen 6 ft. high and 2 ft. deep bordered the rear of the apparatus, preventing *S* from seeing the projector and exploring the rear of the apparatus.

A mirror was mounted above and behind *S*'s chair so that *S*'s performance could be viewed from the adjacent recording room. A child's chair completed the furnishings of the experimental room.

A film strip containing colored pictures of common animals was used to provide visual reinforcement. The film strip contained a total of 35 different animals arranged in two random sequences. Separating each picture was a blank frame. Each picture was illuminated for 1 sec., and between pictures the screen was a uniform gray. The pictures were presented according to a variable ratio schedule with a mean of one picture for every 20 responses.

The presentation of stimuli and recording of responses were controlled by electronic programming and recording equipment located in a room to the rear of the experimental room.

Pretraining task. The pretraining room was 13½ ft. long and 6½ ft. wide and was located next to the experimental room. It contained two tables and four chairs. A one-way vision mirror was located in a door at the rear of the room. On the left wall was a 4 sq. in. speaker and a window covered by a curtain.

For *S*s in the Toy condition the tables and chairs were covered with toys. The toys included a drumming clown, a drinking rabbit, a rocking-telephoning bear, a balloon-blowing bear, a typewriting doll, a bouncing bird, a butterfly-chasing duck, a spiral-shooting gun, a large stuffed bear, a stuffed rabbit, a beach ball, and a xylophone. All except the last seven toys were powered by batteries and were equipped with 2 ft. extension cords equipped with switches which could be easily operated by the children.

Procedure

*S*s were randomly divided into three groups with an equal number of boys and girls in each group. For the two pretraining groups a 15-minute period in the pretraining room preceded the experimental task. In the Toy group the pretraining room contained the toys described above. In the Isolation group no toys were present. The preliminary 15-minute period was omitted for the control group.

*S*s were obtained individually by *E* from their classroom. While *E* escorted *S* to the laboratory room, *E* attempted to gain rapport with *S* by talking about school and related matters.

Pretraining task. For *S*s in the Toy and Isolation groups *E* approached the experimental room, stopped, and told *S*, "Now the game is in that room right there, but I have to get it ready. You wait in here and I will come and get you when the game is ready." *E* walked with *S* into the pretraining room.

In the Isolation condition, *E* told *S*, "As soon as I get the game ready I will come back and get you. You wait in here until I get the game ready." *E* went out of the room leaving the door partly open.

In the Toy condition *E* told *S*, "I have a lot of toys in here and you can play with them while you are waiting. Play with any of the toys you want to and have fun." Again, *E* left the door partly open.

In both the Toy and Isolation conditions instrumental music was played softly through the speaker. This was introduced in an effort to get *S* to relax in the pretraining room and to reduce possible concern *S* might have in playing with the mechanical toys.

As *E* left the pretraining room, he started a stopwatch and went to a hallway from which he could observe *S*'s behavior through the one-way mirror of the pretraining room. *E* was thus able to intercede if S appeared to be disturbed or tried to leave the room during the 15-minute period. Three *S*s, all in the Toy group, were discarded because their pretraining was discontinued when they showed evidence of having to go to the toilet. Three other *S*s had to be reassured once by *E* but then continued pretraining with no further difficulties and were retained as *S*s.

At the conclusion of the 15-minute period *E* returned to the pretraining room and told *S*, "The game is ready. You can come with me now."

Experimental task. *E* took *S* to the experimental room, seated *S* before the apparatus, and said, "This is a game we call the Television Game. See this handle? You pull it down and pictures will come on up here. Now you pull the handle and you can see the pictures." *E* waited until one picture was projected. If *S* stopped pulling the lever before this occurred, *E* urged him to continue pulling. After *S* had seen the first picture, *E* said, "O.K. I would like for you to keep pulling. I have to go to the other room and work the machine. I'll be back when the game is over." *E* then went into the adjacent recording room, where he monitored the electronic apparatus and observed *S*. The experimental task was concluded after seven minutes.

*S*s in the Control group were taken directly to the experimental room, and the procedure for the experimental task was identical to that for the other two groups.

Three *S*s were discarded during their performance on the experimental task; for one *S* the equipment failed, one *S* showed motor impairment, and one *S* verbalized that she thought that she was to remember the name of each picture and responded at a rate slower than any other *S* in the study.

RESULTS AND DISCUSSION

The major results of the study are summarized in Figure 1, which presents the average number of responses made during each minute of the task in each group. The data are presented separately for boys and girls, since preliminary studies with visual reinforcement indicated the possibilities of a sex difference in performance. In both the Isolation and Control groups a rapid increase in the number of responses is evident in the first few minutes; by the seventh minute there is a general decrease in the number of responses. In both groups the number of responses made by boys after the first minute is consistently above that of the girls. In the Toy group performance begins at a high level, continues in a somewhat variable fashion, and begins to decline toward the end of the task. The clear sex differences found in the other two groups do not appear.

The significance of the results can be assessed more adequately by means of an analysis of variance incorporating an analysis of trend (1). This anal-

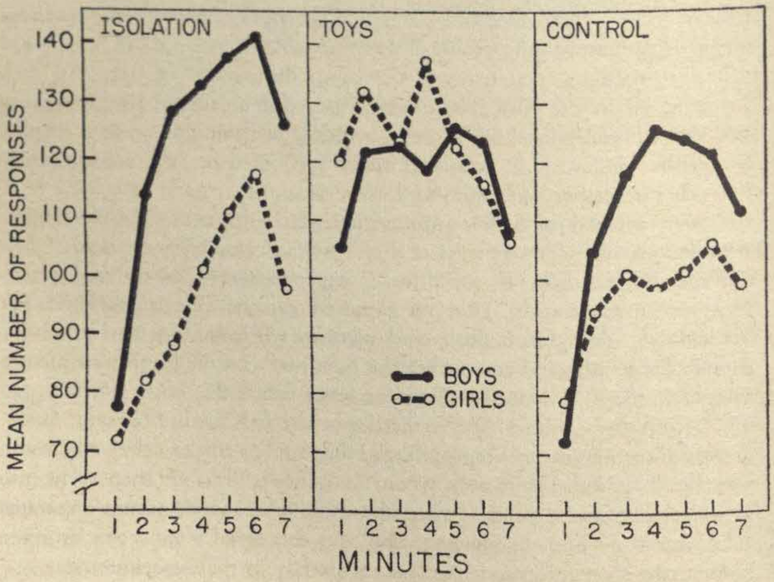

FIGURE 1—Average number of responses made by boys and girls in each group during the seven minutes of the experimental task.

TABLE 1

SUMMARY OF TREND ANALYSIS

Source of Variation	*df*	*MS*	*Error*	*F*
A. Over-all Trend	6	8648.8	E	20.9**
1. Linear	1	21564.5	E.1	17.6**
2. Quadratic	1	27669.1	E.2	48.0**
3. Residual	4	664.7	E.3	3.1*
B. Between Group Means	5	10890.1	D	2.5*
1. Sex (S)	1	13760.0	D	3.2
2. Condition (C)	2	10740.6	D	2.5
3. S × C	2	9604.6	D	2.2
C. Between Group Trends	30	913.3	E	2.2**
1. Linear	5	3652.4	E.1	3.0*
a. Sex	1	1634.3	E.1	1.3
b. Condition	2	8254.7	E.1	6.7**
c. S × C	2	59.3	E.1	. .
2. Quadratic	5	922.4	E.2	1.6
a. Sex	1	2564.3	E.2	4.4*
b. Condition	2	385.0	E.2	. .
c. S × C	2	639.0	E.2	1.1
3. Residual	20	226.2	E.3	1.3
D. Between Individual Means	54	4317.1	E	10.4**
E. Between Individual Trends	324	413.9		
1. Linear	54	1225.0		
2. Quadratic	54	576.7		
3. Residual	216	170.4		

$^{*}p < .05$.
$^{**}p < .01$.

ysis is summarized in Table 1. The linear and quadratic components of trend shown in Figure 1 are highly significant, indicating a general increase in response rate with a curvilinear trend being produced by the decline in response rate toward the end of the seven-minute period.

The *F* related to between-group means is significant, but the over-all performance of boys and girls and the over-all performance in the three conditions do not differ significantly. The effect of each variable on performance is evident, however, when the between-group trends are analyzed. The difference in quadratic trend is significant for the two sexes, and the difference in linear trend is significant for the three conditions. There are significant individual differences in average performance.

The data offer some support for the prediction that a 15-minute period of visual stimulation would reduce the reinforcing effects of visual stimuli. The rapid increase in response rates found in the Isolation and Control groups did not occur in the Toy group. *S*s in the Toy group began response at a high level and maintained this level throughout most of the seven-minute period. There was no indication that visual reinforcement increased the rate of response for these *S*s. The question posed by the data, however, is why the initial rate of response was so high. Two interpretations may be offered: First, the pretraining experience may have increased the strength of the general class of manipulatory responses so that these *S*s entered the experimental task with a strong tendency to respond. A second interpretation is that the pretraining experience had an activating effect which carried over into the experimental task to produce a high level of responding.

Isolation for a 15-minute period did not have a strong effect on performance. The rate of response of boys was consistently higher in the Isolation group than in the Control group, but the rate of response did not differ consistently between the two conditions for girls. One of the problems encountered in this type of study is to provide a meaningful degree of visual deprivation. It was hoped that the lack of novel and interesting objects would result in a higher degree of visual deprivation in the Isolation group than in either of the other groups; however, *S*s had an ample opportunity to explore the room and gain visual experience during the 15 minutes of isolation.

The tendency of boys in the Control and Isolation groups to show a greater increase in response rates than the girls merits further investigation. It is doubtful that the difference is attributable to differences in motor ability, for the initial rates of response were very similar in all cases for boys and girls, and the general level of response of girls in the Toy group was approximately equal to that of the boys. There is the possibility, of course, that these particular pictures were more appealing to boys. Again, this does not seem to be a likely interpretation.

SUMMARY

Three groups of kindergarten and elementary school children were presented with a simple operant task in which lever pressing was reinforced with visual stimuli. A variable-ratio schedule of reinforcement provided a colored picture of a common animal after an average of every 20 responses. A control group and two pretraining groups were included. The first pretraining group played alone for 15 minutes in a room containing a wide variety of interesting toys. The second pretraining group was left alone for 15 minutes with no toys. An analysis of trend of performance over the seven-minute experimental period indicated a significant difference in linear trend associated with condition. The Toy group began response at a higher level than the other two groups, but failed to show the rapid increase in response rate found in the other groups. The difference in the quadratic component of between-group trends associated with sex was significant.

REFERENCES

1. GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, 53, 141-154.

EFFECTS OF PRETRAINING WITH VARIED STIMULI ON CHILDREN'S DISCRIMINATION LEARNING[1]

SHELDON H. WHITE*
University of Chicago

It has been found that a response generalizes more when it has been learned to several stimuli than when trained to a single stimulus (4, 7). If varied training somehow induces *S* to minimize or overlook subsequent stimulus differences in the experimental setting, it should lead to poorer subsequent discrimination learning than equivalent pretraining to respond to a constant stimulus.

The first study to be reported sought to test this hypothesis by pretraining *S*s to respond to stimuli of constant or varied color prior to the learning of a two-choice color discrimination task. One group (C-8) was given eight trials of pretraining to choose the correct stimulus of a subsequent discrimination. A second group (V-24) was given the same training plus equal training to two irrelevant stimuli. A third group (C-24) was given 24 trials of pretraining to choose the correct stimulus.

A second study then sought to determine directly how varied pretraining transfers to discrimination learning, using two control groups. In this study, a group (V-24) given varied pretraining was compared with a group (N-24) given equivalent pretraining to a neutral stimulus and another group (O) which learned the discrimination problem without pretraining.

EXPERIMENT I

Method

Subjects. The *S*s were 60 3½- to 5-year-old preschool children, enrolled in two private Chicago nursery schools.[2] Twenty, 14 from school A and six from school B, were assigned by lot into each of the three experimental

---

* Department of Psychology, University of Chicago, Chicago 37, Illinois.

[1] Experiment 1 was performed with the aid of a Ford Foundation in Behavioral Sciences research grant, experiment 2 with the aid of a research grant (M-3639) from the National Institute of Mental Health, United States Public Health Service.

[2] The author wishes to express his gratitude to Mrs. Marguerite Kelly of the Bo-Peep Nursery School, Mrs. Mannie Tuteur of the Garden Valley Nursery School, Mrs. Dorothy Anker of the Home Club Nursery School, and Miss Jennie Lee Golding and Mrs. Esther Hecht of the South Shore Nursery School, together with their staffs, for their helpful and pleasant cooperation with this study.

Child Develpm., 1961, 32, 745-753.

groups. The children of school A were experimentally naive; about half of the school B children had participated in a previous study which involved discrimination learning.

Apparatus. The stimuli used in pretraining and discrimination were four 3½ in. squares, cut from ⅝ in. plywood and painted white, yellow, red-orange, and red—in Ostwald color notation: a, 3 pa, 6½ pc, and 7½ pc (1). These blocks were always shown on a 9 by 12 in. gray plywood base (Ostwald e). Holes in the base and underside of the blocks enabled a marble to be concealed under a block whether it was placed on the right or left side of the base. A flat-black wooden panel was placed between the child and the base to screen the *E*'s preparation for each trial.

Procedure. At the beginning of his session, the child chose one of a display of small toys to be his prize for playing the game. *E* took the toy and instructed the child that he was to fill the indentations on a marble board in order to win it as a prize.

During pretraining trials, *S* was faced with one block at a time and simply lifted it to find a marble underneath, which he saved towards his toy. *Ss* in group C-8 were given eight pretraining trials to the red-orange block. Group V-24, during 24 trials, had eight trials each to the white, yellow, and red-orange stimuli, with irregular alternation of the stimuli. Group C-24 had 24 trials to the red-orange block. For all groups, successive pretraining stimuli were alternated between the right and left sides in a pattern designed to avoid positional or single-alternation sequences of response.

The following instructions were given between pretraining and discrimination: "Now the game is going to get a little harder. I'm going to show you two blocks at a time, but only one of them is going to have a marble under it. Each time I show you the blocks, you have just one guess to try to find the marble—you can just pick up one block at a time. Try to get a marble each time." Following these instructions, the child was given 40 trials of a two-choice discrimination task, with the red-orange block correct and the red block incorrect. Counterbalancing for position was continued as before, with the correct stimulus appearing equally often on the left and right hand sides.

At the conclusion of the experiment, the child was always awarded his toy and encouraged to feel that he had done well.

Results

Despite the simplicity and repetitiveness of the pretraining procedure, all children responded well during pretraining and, without any obvious flagging of interest, collected their marbles and saved them towards their toy.

Upon transition to the discrimination task, they did not automatically continue with that familiar stimulus which had been bringing them reinforcement. Of the 20 *Ss* in each group, only about half gave a correct response on the first discrimination trial—10 in group C-24, 11 in group

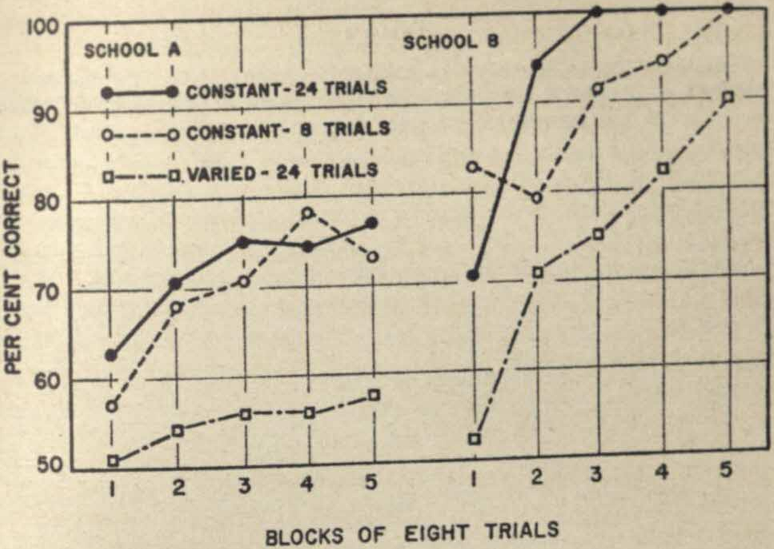

FIGURE 1—Discrimination learning after two degrees of pretraining to choose the correct stimulus (C-24 and C-8) and varied pretraining including the correct and two irrelevant stimuli (V-24). There were 14 children in each of school A groups, and six in each of school B groups (experiment 1).

C-8, and nine in group V-24. In this connection, it is possible that the transitional instructions ("Now the game is going to get a little harder . . .") caused them to take a fresh approach towards the discrimination task.

Complete performances of the three groups on the discrimination task are shown in Figure 1. Despite the fact that the *S*s of school B performed at an appreciably higher level than those of school A, the information conveyed by either set of subgroups appears about the same. All groups showed some tendency to improve their performance over trials. C-24 *S*s performed best, with C-8 *S*s only slightly poorer; the performance of both groups was markedly superior to that of group V-24.

An analysis of variance (2) compared children of different groups and schools for number of correct responses in successive eight-trial blocks of the discrimination learning task (Table 1). The significant Groups effect reflects the superiority of groups C-24 and C-8 over group V-24. These three groups averaged, respectively, 31.4, 30.2, and 24.3 correct responses in 40 trials. Comparisons by t tests showed group V-24's performance to be significantly poorer than those of groups C-24 and C-8 ($p < .01$ and $p < .05$); the C-24 and C-8 groups did not differ significantly from each other ($t < 1$).

The significant Schools effect shows that the superior performance of *S*s in school B is statistically reliable. The significant Blocks effect reflects the increase over trials shown by all subgroups. The two interactions which

TABLE 1

SUMMARY OF EXPERIMENT 1—ANALYSIS OF VARIANCE COMPARING NUMBER OF CORRECT RESPONSES IN EIGHT-TRIAL BLOCKS OF CHILDREN OF DIFFERENT GROUPS AND SCHOOLS

Source	*df*	*MS*	*F*	*p*
Between Subjects	59			
Groups (G)	2	57.22	5.22	.01
Schools (S)	1	161.60	14.74	.001
G × S	2	0.09	< 1.00	
error (b)	54	10.96		
Within Subjects	240			
Blocks (B)	4	17.99	12.94	.001
G × B	8	0.44	< 1.00	
S × B	4	2.86	2.06	.10
G × S × B	8	1.76	1.27	.10
error (w)	216	1.39		
Total	299			

approach significance in the table probably derive from the following trends (Figure 1): *S*s in school B showed a steeper increase over trials than those in school A, which contrast between schools is relatively more marked for group V-24 than for the other two groups.

EXPERIMENT 2[3]

The first experiment showed that varied pretraining leads to poorer discrimination learning than constant pretraining. Is the varied pretraining less facilitating than the direct pretraining, or does it actually interfere with a subsequent discrimination? The second experiment sought an answer to this question by a direct assessment of the effects of varied pretraining on discrimination.

Method

The *S*s were 60 3½- to 5-year-old children, experimentally naive, drawn from two additional Chicago nursery schools. Twenty, 10 from school C and 10 from school D, were assigned by lot into each of the three experimental groups.

The varied pretraining of group V-24 has been described above. Group O had no pretraining and was intended as an absolute control, whose comparison with group V-24 would show whether varied pretraining has a net effect of facilitating or interfering with discrimination learning. A second control group was considered necessary because of the generally facilitative effect which *any* sort of pretraining might be expected to produce, by "warm-up." Thus, group N-24 received 24 trials of pretraining to a neutral, gray (Ostwald 15 ml) block before going on to the discrimination. It was intended as a relative control, whose comparison with group V-24 would show whether varied pretraining produces more or less facilitation than can be attributed to its warming-up effect.

---

[3] Gratitude is here expressed to Mr. Paul Grim, who collected the data and assisted in its analysis for this experiment.

The discrimination task and other aspects of the method and procedure were as described above for experiment 1.

Results

In this experiment, the first trial of the discrimination showed choices biased towards the red-orange, correct stimulus. Forty-seven of the 60 *Ss* chose it, a proportion far beyond chance (binomial $p < .0001$). Of the 47, 14 were in group V-24, 16 in group N-24, and 17 in group O. On the next three trials, the red-orange stimulus was chosen 26, 21, and 36 times out of a possible 60. The most plausible interpretation of this choice pattern was in terms of the standard procedure of stimulus placement, with a right-position preference strong on the first trial but fading over the next few trials.

Table 2 shows the mean number of correct responses over 40 trials for each of the three groups. Although there are apparent differences among the groups—principally, a greater number of correct responses for groups N-24 and V-24 than for group O—these differences were not confirmed in over-all analysis of the experiment 2 data. Analysis of variance with Groups, Schools, and Blocks (of eight trials) as factors showed only one significant entry: the Blocks effect, reflecting a reliable general increase in performance over successive blocks of trials.

TABLE 2

DISCRIMINATION PERFORMANCE (MEAN NUMBER CORRECT IN 40 TRIALS) FOR THE THREE GROUPS OF EXPERIMENT 2

	Group N-24	*Group V-24*	*Group O*	*All Groups*
School C	27.7	24.8	22.9	25.1
School D	26.1	26.5	21.1	24.6
Both Schools	26.9	25.6	22.0	24.8

COMBINED ANALYSIS OF EXPERIMENTS 1 AND 2

In each of the two experiments, the replications in different schools (school A vs. school B in experiment 1; school C vs. school D in experiment 2) generally confirmed each other, but were by no means numerically identical in result. As was indicated by the significant Schools effect in experiment 1, such differences between replications do not necessarily come from sampling error but may reflect genuine population differences. For this reason, some caution must be exercised in comparing experimental groups drawn from different sample sources.

Some justification for cross-comparisons between experiments 1 and 2 may be found in the fact that their identically-treated V-24 groups gave mean performances, respectively, of 24.3 and 25.6 correct responses in 40 trials—with the difference between them being statistically and practically not significant. For this reason, and because there is a degree of "face validity" about the results, the cross-experiment comparisons are to be described. Conclusions arising from these comparisons depend upon the comparability of the experimental samples, and this reservation must be emphasized.

The first column of Table 3 shows the means of the five different groups obtained in the two experiments (with both V-24 samples pooled). A simple analysis of variance, comparing the five groups, yielded an F of 5.12 (4,115 *df*), significant at the .001 level of confidence.

TABLE 3

PERFORMANCE OF THE TREATMENT GROUPS OF EXPERIMENTS 1 AND 2 AND *t* TEST COMPARISONS

				t TEST COMPARISONS			
Group	*Mean*	*SD*	*N*	*C-24*	*C-8*	*N-24*	*V-24*
C-24	31.4	8.57	20				
C-8	30.2	9.19	20	0.41			
N-24	26.9	8.93	20	1.61	1.15		
V-24	24.9	7.41	40	2.87**	2.24*	0.86	
O	22.0	3.96	20	4.43**	3.66**	2.24*	1.97

* Significant at .05 level.

** Significant at .01 level.

The right half of Table 3 shows the results of *t* test comparisons among the groups. Because the variance of the group O scores was markedly less than the variance of the other groups, all significance levels are for degrees of freedom determined by a formula which takes into account inequality of variances (6).

As one might expect, direct training to the correct stimulus is followed by the best discrimination performance, although it is perhaps somewhat surprising that 24 trials of such pretraining has little more effect than does eight trials. There is a noticeable superiority of this sort of training over a simple "warming-up" sort of pretraining (group N-24); the statistics do not show this superiority to be reliable, but, nevertheless, the hypothesis probably might still be entertained that the facilitating effects of C-24 and C-8 pretraining are due to something more than simple warm-up.

That warm-up itself facilitates discrimination learning is shown by the significant superiority of group N-24 to group O. Group V-24 lies between groups N-24 and O, superior to O by a *t* significant only at the .10 level of confidence. Consequently, the interpretation of this kind of pretraining must be reserved; it probably is facilitative, although perhaps less than one might expect from its warm-up effect alone. Varied pretraining is certainly less facilitating than direct pretraining, even to that amount of direct pretraining (eight trials) which it itself includes.

DISCUSSION

In experiment 1, the rather surprising finding that group C-24 was not markedly superior to group C-8 was noted, implying that 24 trials of direct pretraining to the correct stimulus is hardly more helpful to a subsequent discrimination than only eight such trials. Since the performance of both groups was at a rather high level, there is the possibility that group C-24 was hindered by a "ceiling effect" from expressing a potentially stronger degreee of superiority. This possibility cannot be fully discounted; however, it will be noted in Figure 1 that the C-24 – C-8 difference was about the same for schools A and B (quantitatively, about 4 per cent in each school) despite the fact that school B *S*s were visibly operating much closer to ceiling than school A *S*s. There may, really, be little difference in the effectiveness of these two degrees of direct pretraining.

The differences between the schools of experiment 1 may have come from several causes. Some of school B *S*'s were participants in an earlier discrimination study, and Shepard (5) has demonstrated a learning-to-learn effect for children's discriminations which is particularly potent from the first to the second problem. School B staff participated more, and was more supportive, in the recruitment of children for the experiment. Finally, there is the possibility of a relatively subtle socioeconomic difference: both schools enroll their children from substantially the same neighborhood, but school B is reputed to attract relatively more professional and educated people.

It was the author's impression that the two nursery schools of the second experiment differed more in character of administration and socioeconomic sample than did those of experiment 1, yet no strong difference between schools C and D was found.

Group N-24 of the second experiment requires some interpretation. As noted above, it was included as a measure of the general warming-up effects of single-stimulus pretraining. Facilitating transfer from warm-up is a well-substantiated effect in adult human learning (3) and is probably equally effective in children's learning, although there does not appear to be a direct experimental demonstration to this effect. Its effects upon adult learning have been attributed to set factors, particularly "attentional" and postural adjustments to the experimental task. In the case of children, particularly where they are experimentally naive, it seems probable that there is an additional, emotional component of the child's warming-up to the situation —the overcoming of shyness before a relatively strange adult and his relatively strange new task. It might be an interesting question for further study to examine whether sophisticated preschool children would show quite as much benefit from pretraining as do naive children.

It appears that four of the pretraining groups—C-24, C-8, N-24, and O—arrange themselves in an order which is quite consistent with expectations from established mechanisms. Presumably the two best-performing groups would be superior through a combination of associative transfer of the habit of responding to the red-orange stimulus, plus nonspecific facilitating transfer from warmup. The lesser facilitation produced by N-24 pretraining would be based on warmup alone.

Group V-24, like group C-8, had eight trials of training to respond to the subsequently correct stimulus, but its additional training to irrelevant stimuli was apparently deleterious. Apparently the direct training is "lost" amongst the irrelevant. The performance of this group was below what one might have expected by warm-up alone, though not reliably so.

That varied pretraining has produced more stimulus generalization than constant pretraining is explainable either by the greater resistance to extinction which some studies have found after varied pretraining, or else by a "generalization set" (7). Since the influence of resistance to extinction is not so easily distinguishable in the present discrimination paradigm as it is in most generalization procedures, it is tempting to explain these results by the generalization set mechanism—with the varied pretraining serving as an "implicit instruction" that the response is to be widely attached or that stimulus differences may safely be overlooked.

SUMMARY

In experiment 1, three groups of 20 preschool children each, drawn from two nursery schools, were given a pretraining procedure in which they saved marbles found under a single colored block. Group C-24 was presented a block of red-orange color 24 times, group C-8 the red-orange block eight times. Group V-24 was presented blocks of varied color—the red-orange, a yellow, and a white block—eight times each. Following their pretraining, all *S*s were given 40 trials of a simultaneous color discrimination with the red-orange block correct and a red block incorrect. Groups C-24 and C-8 did not differ significantly from each other in discrimination performance, but were significantly superior to group V-24. Although the level of discrimination performance differed significantly between nursery schools, the essential experimental comparisons between groups were the same for both schools.

In experiment 2, there were also three pretraining groups of 20 preschool children each, who subsequently learned the same discrimination. In addition to group V-24, described above, group N-24 had 24 trials of pretraining to a neutral gray stimulus, and group O had no pretraining. Group N-24 was significantly superior to group O in discrimination learning, with group V-24 in between, not significantly different from either. Although the children were drawn from two nurseries, there were no schools differences.

Cross-comparisons between the two experiments ranked the five treatments in the following order in promoting discrimination: C-24, C-8, N-24, V-24, and O.

REFERENCES

1. JACOBSON, E., GRANVILLE, W. C., & FOSS, C. E. *Color harmony manual.* (3d Ed.) Container Corp. of America, 1948.
2. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Houghton Mifflin, 1953.
3. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning.* Longmans, Green, 1952.
4. RIOPELLE, A. J., & ROGERS, J. P. Transfer of training to orthogonal dimensions. *J. exp. Psychol.,* 1956, 52, 367-370.
5. SHEPARD, W. O. Learning set in preschool children. *J. comp. physiol. Psychol.,* 1957, 50, 15-17.
6. WALKER, H. M., & LEV, J. *Statistical inference.* Holt, 1953.
7. WHITE, S. H., & SPIKER, C. C. The effects of a variable conditioned stimulus upon the generalization of an instrumental response. *Child Develpm.,* 1960, 31, 313-319.

SECONDARY REINFORCEMENT WITH PRESCHOOL CHILDREN[1,2]

JANE G. FORT*

University of Massachusetts

Although there is abundant literature in the area of secondary reinforcement, little, if any, of it can be considered conclusive evidence that stimuli can become and remain strong secondary reinforcers. In a recent review of the literature on secondary reinforcement (13), contradictory results obtained by Melching (12) and by N. A. Myers (14) were pointed out. The resolution of this difference in results could be a significant step in clarifying findings and their interpretations.

Melching (12) attempted to find secondary reinforcement effects in bar-pressing behavior subsequent to 50 per cent neutral stimulus (buzz) presentation in training. One half of these *Ss* underwent 100 per cent buzz in the extinction period, while the other half was presented no stimulus. The results indicated no significant difference in number of extinction responses made by these groups. Melching explains his results through stimulus generalization. Since the percentage of responses on which the buzz was experienced in training was changed an equal amount for both groups, equivalent extinction responses could then be expected.

Myers (14) used a controlled operant situation to study partial and regular reinforcement. Children were trained to make a series of responses

---

* Department of Psychology, University of Massachusetts, Amherst.

[1] This research was supported by funds provided by NIMH Grant M-2620.

[2] This article is based on a thesis submitted in partial fulfillment of the requirements for the Master's degree, Department of Psychology, University of Massachusetts, 1960 (6). The author wishes to extend her appreciation to her committee members, Drs. R. E. Wyman, A. E. Goss, N. A. Myers, and especially J. L. Myers for their patience, cooperation, and helpful advice.

Child Develpm., 1961, 32, 755-764.

for a candy reward. Contrary to Melching's findings, Myers found that, for each of her groups trained identically, the one given the newly acquired secondary reinforcer (chip) in extinction consistently made more responses in extinction than the one given no chip.

The discrepancy between these results and those of Melching may be due to one or a combination of factors. Melching employed rats in a free operant situation, using, for the potential secondary reinforcer, a buzz presented by the *E* at the same time as the primary reward. Myers' children had to manipulate a chip in order to receive their primary reward. It is possible that this handling enabled the chip to acquire greater reward value than did the mere hearing of the buzz used in Melching's study. However, species differences could be great enough to cause the cited discrepancy. The present study, therefore, was primarily designed to help clarify the results of the Melching and Myers studies. In addition, data were provided on partial primary reinforcement effects with children.

METHOD

Subjects

The *Ss* were 144 preschool children ranging in age from 3 years to 5 years, 11 months.[3] Within the restrictions of equating the groups for age and for sex, the *Ss* were equally and randomly divided among the 12 groups.

Apparatus

The apparatus, designed to attract and maintain the interest of preschool children, was a clown face, having red jewel-light eyes, a push-button nose, and a slot-tray mouth. Candy[4] (primary reinforcer) was dispensed through a tube to the mouth of the clown, while a 1-second buzz (potential secondary reinforcer) was heard from the interior of the clown. *E* had access to and operated two silent knife switches which allowed administration of the predetermined reinforcement. The number of responses was recorded on an electric magnetic counter mounted on the back of the clown, out of *S*'s sight. *E* recorded the cumulative number of responses made during each successive 30 seconds of extinction.

Procedure

A room separate from general distractions was chosen for the experimental session. The clown was placed on a small table or on the floor so that *S* had easy access to the nose. Subsequent to miscellaneous conversation designed to reduce possible tension, *S* accompanied *E* into the room. Attention was called to the clown face and especially to the nose. *S* was told that "something funny happens when you press his nose. Let's see what happens." *E* pressed the clown's nose and thereby received a candy. *S* was encouraged to try it also and was given one rewarded preliminary trial. He was then told that he could stay and continue to play the game "as long as you want." *S* and *E* remained alone in the room throughout the experimental session.

---

[3] The author extends her appreciation to the many Amherst, Massachusetts nursery schools and kindergartens and in particular to William Barry, Superintendent of Schools, and Miss Esther Wien, Elementary School Supervisor, Northampton, Massachusetts, for their cooperation in providing subjects and facilities for this study.

[4] The candies are manufactured by the Kirkman Pharmacal Company of Seattle, Washington. They contain no medication or sugar, but are sweet and come in four assorted colors and flavors: lime, lemon, orange, and raspberry.

Each *S* was allowed 20 responses during training. According to a random and predetermined schedule for each group, *S* could experience one of four treatments following each response: (a) a candy alone, (b) a buzz alone, (c) a candy and a buzz, or (d) neither a candy nor a buzz. For each group, candy was presented with or without a buzz on the first and the last responses in training. On the next response after the 20 training responses were completed, the extinction period was noiselessly begun without comment from *E*. Each *S* was run until he stopped and indicated a desire to leave the room. If *S* had not stopped by the end of 10 minutes in extinction, *E* terminated the session. Questions and comments were ignored as much as possible, but if necessary, *E* made a very noncommittal response. To reduce frustration which might arise from the absence of the primary reward in extinction, each *S* was given one last candy at the end of the extinction period.

Experimenal Design

During training, two schedules of primary reinforcement (100 per cent and 50 per cent candy) were used along with three schedules of potential secondary reinforcement (100 per cent, 50 per cent, and 0 per cent buzz). During extinction, two schedules of buzz (100 per cent and 0 per cent) were employed. There were 12 *Ss* in each of the resulting 12 groups.

RESULTS

To reduce heterogeneity of variance, square root transformations were applied to the data. Analysis of variance was then run on the transformed total number of responses in extinction. Table 1 shows a summary of the analysis. Only the partial primary reinforcement effect was significant. The *F* was 15.05, yielding significance beyond the .001 level. Those *S*s who chose to stop before the end of 10 minutes were given values of 0 for the remaining periods of time. Figure 1 shows the subsequent mean number of responses in extinction over 10 one-minute periods of time when the two percentages of primary reinforcement (100 per cent and 50 per cent) are plotted. The curves in both cases show a steep initial drop within the first two minutes and a subsequent gradual decrease to the same level. This figure gives a strong indication of the partial primary reinforcement effect. The groups which received 50 per cent candy presentation in training made a mean of 78 responses in the first minutes of extinction, while those trained

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE

Source	*df*	*MS*	*F*
Total	143		
Primary Reinforcement (P)	1	236.44	15.05**
Secondary Reinforcement (S)	2	2.52	.16
Extinction Reinforcement (E)	1	1.56	.09
P × S	2	20.52	1.30
P × E	1	28.62	1.82
S × E	2	11.99	.76
P × S × E	2	3.44	.21
Ss/P × S × E	132	15.71*	

* Error term for all sources of variance.

** $p < .001$.

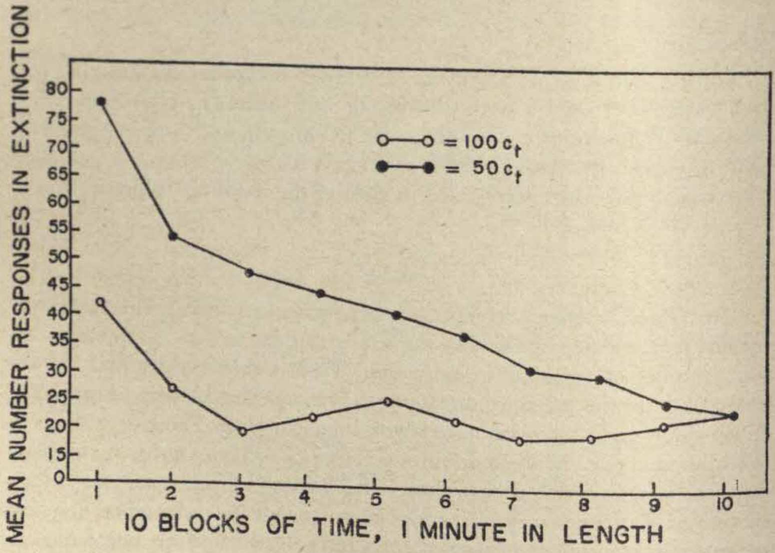

FIGURE 1—Mean number of responses in extinction in each of 10 one-minute periods of time for groups receiving 100 per cent and 50 per cent primary reinforcement in training.

TABLE 2

SUMMARY OF RESULTS FOR NUMBER OF RESPONSES IN EXTINCTION*

	TRAINING CONDITION		EXTINCTION CONDITION	*Total Number Responses in Extinction*
Group Number	*Candy—%*	*Buzz—%*	*Buzz—%*	
9	50	50	100	1174
11	50	0	100	1153
2	100	100	0	1015
10	50	50	0	843
8	50	100	0	833
7	50	100	100	657
12	50	0	0	589
6	100	0	0	522
1	100	100	100	480
5	100	0	100	440
4	100	50	0	331
3	100	50	100	307

* The groups may be referred to by the respective percentages of reinforcement administered in training and extinction; e.g., 100-50-0 denotes the group which received 100 per cent candy and 50 per cent buzz in training, and was presented 0 per cent buzz in extinction.

on 100 per cent primary reinforcement averaged 44 responses during this period.

Table 2 shows the total number of responses made by each group in extinction with the respective training and extinction conditions indicated. The groups are rank-ordered from the largest number of responses to the smallest. Here, it can be seen that with one exception (group 2: 100-100-0), the groups trained under the partial primary reinforcement schedule rank above those trained under the 100 per cent reinforcement schedule. The column which notes the percentage of responses on which the buzz was presented in training shows that the groups are not orderly with respect to this variable, but seem to be randomly distributed. Although the highest group in number of extinction responses was one which received the buzz on 50 per cent of the responses in training, two other groups which were also trained under this schedule (group 3: 100-50-100, and group 4: 100-50-0) were the lowest groups of all in number of responses in extinction. It may be noticed that the second highest group was one having never experienced the buzz in training and had it introduced in extinction (group 11: 50-0-100). The same random distribution noted in training is also noted for the percentage of responses on which the buzz was presented in extinction. Aside from the fact that the two highest groups were presented 100 per cent buzz in extinction, there is no regularity shown for this variable. Table 3

TABLE 3

SUMMARY OF RESULTS FOR MEANS AND VARIANCES

	CONDITIONS DURING EXTINCTION			
	100% Buzz		0% Buzz	
Conditions During Training	*Means*	*Variances*	*Means*	*Variances*
100% *Candy During Training*				
100% buzz	40.0	3,506.0	84.6	15,546.0
50% buzz	25.6	2,469.0	27.6	1,026.0
0% buzz	36.7	2,780.0	43.5	2,298.0
50% *Candy During Training*				
100% buzz	54.8	1,123.0	69.4	2,298.0
50% buzz	97.8	11,924.0	70.2	2,554.0
0% buzz	96.1	8,750.0	49.1	1,737.0

shows the means and variances for the 12 groups. Here, the heterogeneity of variance can easily be seen.

A Duncan range test (4) performed on the mean number of responses for each group showed that the two highest groups in number of extinction responses were significantly different from the two lowest groups. Group 9 (50-50-100) and group 11 made significantly more responses in extinction than did either group 3 or group 4.

DISCUSSION

In most studies of partial primary reinforcement, *Ss* have been rats or adults (8, 10); children have been used in only a few. Lewis (9) used 6- and 7-year-old boys in a gambling situation with toys as the primary reward. *Ss* were told that they could win more toys than they had been given, but that they could also lose all of those given them. Partial reinforcement percentages of 50 and 60 per cent produced greater resistance to extinction than 100 and 0 per cent reinforcement. Fattu, Mech, and Auble (5) trained 3- to 5-year-old children to pull a lever for candy. They found a group given 25 per cent reinforcement made more responses during 14 minutes of extinction and took longer to extinguish than did either 50 or 100 per cent reinforcement groups. Myers (14) also found a significantly greater number of responses in extinction following 50 per cent than following 100 per cent reinforcement. Thus, the results of the present study agree with those of other studies in which *Ss* were children.

In view of results found with partial primary reinforcement, one might anticipate that partial secondary reinforcement groups would exhibit a similar effect. However, there was no indication of a secondary reinforcement effect in the present study. Number of responses made in extinction was not related to percentage of secondary reward in training. It was also noted that *Ss* presented the buzz on all responses in extinction did not make significantly more responses than those to whom the buzz was not presented.

The absence of a secondary reinforcement effect in the present study may have been due to several factors. First, there is the possibility of aversive properties of the buzz. However, if the buzz is a noxious stimulus, the groups which did not experience the buzz during training and then had it presented in extinction should have made fewer responses in extinction than those which did experience the buzz in training. This, however, is not the case, for the second highest group (group 11: 50-0-100) in terms of number of responses in extinction is a group which never experienced the buzz in training, but which had it suddenly introduced with the onset of extinction. Although the buzz may not be an aversive stimulus, it has been pointed out elsewhere (13) that stimuli are not really "neutral" and may not have equal initial reinforcing properties. The results of the present study would tend to support this idea. Some *Ss* in this experiment were trained under experimental conditions similar to those used by Myers (14). One difference is that the stimulus employed as the potential secondary reinforcer was a buzz in the present study, but a chip in the Myers study. It would seem from comparisons of groups which received similar treatments in these two studies that the chip has greater reinforcing potential than does the buzz.

Second, there is the possibility that the candy used in this study was not sufficiently reinforcing in and of itself to enhance the value of the buzz. Wyckoff (17) has suggested the need for very strong primary reinforcers in order to obtain an appreciable secondary reinforcement effect. However, there is no particular reason to doubt the rewarding properties of the candy; no *Ss* reported distaste for it, and it has been previously used successfully (2). There does remain a definite need for thorough research on various kinds of reinforcers and evaluation of their effectiveness. Valuable information may well be gained from actual scaling of stimuli proposed for use as reinforcers for children.

Third, the absence of a secondary reinforcement effect could be a result of too few pairings of the primary reinforcer and the potential secondary reinforcer. A few studies have investigated the effects of number of pairings of primary reinforcement and neutral stimulus on secondary reinforcement. Bersh (1) studied varying numbers of primary and neutral stimulus pairings and found that only his 20- and 120-pairing groups differed significantly from each other. This same general tendency was found by Hall (7), whose 75-reinforcement group was significantly different from his 25-reinforcement group. The Melching study found secondary reinforcement effects with 160 training trials. However, with children, Myers found that as few as 10 trials were sufficient to yield a significant secondary reinforcement effect.

Fourth, there is the possibility that the lack of a secondary reinforcement effect is the result of a great deal of *S* variability. It may be that much of the variance within groups is due to wide variation between *S*s in initial free operant levels. This, however, is doubtful, for the variation in the present study is not much different from that in a study in which strong secondary reinforcement effects was found (14). In addition, the ordering of the groups is so unsystematic as to suggest that the effect, if existing at all, is extremely slight and difficult to detect in situations employing children as *S*s and a buzz as the potential secondary reinforcer. However, the use of a preliminary measure by which pretraining operant level may be controlled for each treatment group should enable a clearer demonstration of the effects of the particular treatments employed.

None of the above factors, alone, then, seems sufficient to account for the lack of a secondary reinforcement effect in the present study. Very few studies have shown durable secondary reinforcement effects, and these studies have either employed special training schedules (e.g., 11, 15, 18) or a manipulable stimulus (e.g., 3, 14, 16). The previously-mentioned discrepancy, then, between the Melching and Myers studies would seem to be explained by differences in treatment of the potential secondary reinforcer. The children in the Myers study had to handle the chip and use it to receive the primary reinforcement, whereas Melching's rats had no physical contact with the buzz. It would seem that this physical contact with the chips led to much greater enhancement of their original value than was true of Melching's buzz or of the buzz in the present study. The results of the present study, then, would tend to indicate that mere pairing of a stimulus with a primary reinforcer is *not* a sufficient condition to establish noticeable and durable secondary reinforcement effects and that the mode of presentation and treatment of the potential reinforcer is of crucial importance.

SUMMARY

The present study was concerned with reinforcement and extinction of a free operant response in preschool children. One hundred and forty-four *S*s were individually given an opportunity to "play a game" with a clown. When *S* pressed the nose, a candy fell into the clown's mouth. At the same time, a buzz was heard from his interior. Twelve experimental groups were used with both partial and regular schedules for the candy and the buzz. *S*s were given the predetermined reinforcement on the first 20 responses, which constituted the training period. Schedules of 100 per cent and 50 per cent candy and 100 per cent, 50 per cent, and 0 per cent buzz were used. In the extinction period, no candy was given and the buzz was presented on 100 per cent or 0 per cent of the responses until *S* indicated a desire to end the session. Otherwise, *E* terminated the session at the end of 10 minutes.

For each *S* the total number of responses in the extinction period was recorded. The results of this study indicated that:

1. Groups trained with 50 per cent candy reinforcement made more responses in the extinction period than did those trained with 100 per cent candy reinforcement.

2. There were no significant differences between groups as a result of the percentage of responses on which the buzz was presented in training or extinction. That is, there was no significant secondary reinforcement effect. Several possible explanations for this were discussed. It was concluded that treatment of the potential secondary reinforcer is important in the demonstration of noticeable and durable secondary reinforcement effects.

REFERENCES

1. BERSH, P. J. The influence of two variables upon the establishment of a secondary reinforcer for operant responses. *J. exp. Psychol.*, 1951, 41, 62-73.
2. BIJOU, S. W. A standard "candy" for research with children. *J. exp. Anal. Behav.*, 1958, 1, 314.
3. COWLES, J. T. Food-tokens as incentives for learning by chimpanzees. *Comp. Psychol. Monogr.*, 1937, 14, No. 5.
4. DUNCAN, D. B. A significance test for differences between ranked treatments in an analysis of variance. *Va. J. Sci.*, 1951, 2, 171-189.
5. FATTU, N. A., MECH, E. U., & AUBLE, D. Partial reinforcement related to "free" responding in extinction with pre-school children. *J. exp. Educ.*, 1955, 23, 365-368.
6. FORT, J. G. Extinction of a free-operant response in pre-school children following partial and regular schedules of primary and potential secondary reinforcement. Univer. of Mass. Library, 1960.
7. HALL, J. F. Studies in secondary reinforcement: I. Secondary reinforcement as a function of the frequency of primary reinforcement. *J. comp. physiol. Psychol.*, 1951, 44, 246-251.
8. JENKINS, W. O., & STANLEY, J. C., JR. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, 47, 193-234.
9. LEWIS, D. J. Partial reinforcement in a gambling situation. *J. exp. Psychol.*, 1952, 43, 447-450.
10. LEWIS, D. J. Partial reinforcement: a selective review of the literature since 1950. *Psychol. Bull.*, 1960, 57, 1-29.
11. MCCLELLAND, D. C., & MCGOWN, D. R. The effect of variable food reinforcement on the strength of a secondary reward. *J. comp. physiol. Psychol.*, 1953, 46, 80-86.

12. MELCHING, W. H. The acquired reward value of an intermittently presented neutral stimulus. *J. comp. physiol. Psychol.*, 1954, 47, 370-374.

13. MYERS, J. L. Secondary reinforcement: a review of recent experimentation. *Psychol. Bull.*, 1958, 55, 284-301.

14. MYERS, N. A. Extinction following partial and continuous primary and secondary reinforcement. *J. exp. Psychol.*, 1960, 60, 172-179.

15. SALTZMAN, I. J. Maze learning in the absence of primary reinforcement: a study of secondary reinforcement. *J. comp. physiol. Psychol.*, 1949, 42, 161-173.

16. WOLFE, J. B. Effectiveness of token rewards for chimpanzees. *Comp. Psychol. Monogr.*, 1936, 12, No. 60.

17. WYCKOFF, L. B. Toward a quantitative theory of secondary reinforcement. *Psychol. Rev.*, 1959, 66, 68-78.

18. ZIMMERMAN, D. W. Durable secondary reinforcement: method and theory. *Psychol. Rev.*, 1957, 64, 373-383.

SECONDARY REINFORCEMENT AS A FUNCTION OF THE NUMBER OF REINFORCED TRIALS[1]

NANCY A. MYERS,* GRACE J. CRAIG,
and JEROME L. MYERS

University of Massachusetts

The purpose of this study was to observe the effects of varying habit strength sustained through secondary and primary reinforcers upon extinction behavior of children in an operant situation. A few studies (1, 3, 4, 6) have investigated the relation between number of training trials and extinction with secondary reinforcement of rats. Little research has been concerned with children's habit strength behavior. Lewis and Duncan (5) investigated the resistance to extinction of a lever-pulling response of college students after 3, 6, 12, or 21 acquisition plays with either 33 per cent, 67 per cent, or 100 per cent reward. Their results corroborated those of Capaldi with rats, indicating that the larger number of acquisition trials resulted in fewer plays to extinction. Siegel and Foshee (8) studied the law of primary reinforcement with preschool children, using 2, 4, 8, or 16 candy reinforcements for bar-pressing, and reported a positive relation between resistance to extinction and number of prior reinforcements. Spiker (9), however, found no significant increase in mean number of lever-pulling responses after 24 reinforcements than after 12 with preschool children, and Myers (7) found no differences in children's extinction behavior after 10 or 20 preceding candy reinforcements.

This study will evaluate the effects of number of training trials, with and without secondary reinforcement, on behavior of preschool children extin-

---

* Department of Psychology, University of Massachusetts, Amherst.

[1] This research was supported by funds from an NIMH Grant, M-2620.

Child Develpm., 1961, 32, 765-772.

guished with and without secondary reinforcement. Thus, it should help resolve the results of studies with rats and with humans, as well as extend the data on secondary reinforcement effects.

METHOD

Apparatus

The apparatus employed was a portable box designed to have eye-appeal and interest for children 3 to 6 years of age. It was placed on a low kindergarten table, and the child sat before it on a low chair. The front of the box was painted with the face of a clown. The back of the box was open for easy access to the candy and token reinforcement mechanisms, two counters, and three switches, by means of which *E* could turn *off* or *on* the lights, the token reinforcer, and the candy reinforcer. The clown's eyes were 6-watt bulbs which lit up at the start of each session and remained on throughout the session. The nose was a red cylinder protruding from the box, and on its end were two small push-buttons mounted horizontally parallel and an inch apart. Both the candy reinforcer and the token reinforcer delivered their contents to the mouth of the clown. A slot for inserting tokens was cut in one cheek, and a small box beneath this, on the inside of the apparatus, collected the tokens.

The token and candy dispensers were identical units differing only in the diameter of the loading tubes. They were designed to release an individual token or candy, respectively, to a slanted metal slide down which it traveled to the mouth of the clown. The loading tube for tokens permitted the stacking of 200 tokens; that for candy was tall enough to accommodate 30 candies. These tubes were refilled before each experimental session. The tokens used were small poker chips, about one half standard size and cut from four colors of plastic: red, white, blue, and yellow. These colors appeared randomly. The candy reinforcements were small chocolates with different colored shells, the popular M & M chocolates produced by Marlon & Marlon, New York.

Subjects

The *S*s were 48 boys and 48 girls between the ages of 5 years, 1 month, and 6 years, 4 months, attending kindergarten classes in the Easthampton, Massachusetts, Schools.[2]

Instructions

No written instructions were read to *S*, but *E* standardized the verbal instructions as much as possible. Each *S* was asked in the classroom if he would like to play a clown game. If *S* said yes, the teacher's permission was obtained to leave the room with *E*. Once in the testing room *S* was instructed to sit at the small chair in front of the small table holding the box, and *E* would show him how to play the game.

---

[2] The authors wish to thank S. J. Bernard, Superintendent of Schools in Easthampton, Massachusetts, and his principals and teaching staff for their cooperation in permitting us access to the students and the school facilities.

The *S* was instructed through the following sequence: press the left push-button on the nose, receive a token, deliver the token to the slot in the cheek, push the right push-button on the nose, receive and eat the candy. This sequence was sufficient to establish the response in all *Ss*. *E* then sat behind the box on the other side of the table, with easy access to the switches. As soon as the eyes were illuminated to signal the start of the game, *S* pressed the first button, and conditioning trials ensued. The eye-lights remained on continuously throughout conditioning and extinction. *Ss* were told they could eat the candies, and in most cases consumed them at once. The only variations in the technique were for the four control groups (P-O groups) which never received tokens. They were instructed in the same manner, but were told that one button was not working, to push the second button only, and then receive a candy directly.

Design

All *Ss* (except P-O groups *Ss*) received one initial trial on the basic response sequence, in which they observed the lights, pressed the first nose-button, received a token, delivered it to the slot in the cheek, pushed the other nose-button, and received a candy. They were then assigned to the eight groups of eight *Ss* each, four boys, four girls, for the remaining reinforcement and extinction procedures. The groups were equated for mean age. They differed in number of training trials and in extinction conditions. Two groups received two training trials, two groups received four training trials, two groups received eight trials, and two groups received 16 trials before extinction. One of each of the groups at each habit-strength value was then extinguished with *tokens* delivered when the token button was pressed (S^r-S^r groups); the other received no tokens for token-button pressing (S^r-O groups). No groups received candy for pressing the candy button in extinction. Two counters reported the cumulative number of responses made on each button, and *E* recorded the number of responses made on each button in each minute of the 5-minute extinction period.

P-O groups for which the token dispenser was not used during training received 2, 4, 8, or 16 trials with candy reinforcement and no candy for button-pressing in extinction.

RESULTS

An analysis of variance was performed on the square-root transforms of the data of the eight experimental groups with respect to total number of extinction responses on both buttons. A summary of this analysis may be seen in Table 1. There were no differences in total number of extinction responses as a function of number of training trials. There were also no

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE PERFORMED ON TOTAL RESPONSES ON BOTH BUTTONS FOR EIGHT GROUPS (S^r-S^r AND S^r-O)

Source	*df*	*MS*	*F*
Number of Trials (Tr)	3	.98	.27
S^r	1	9.03	2.49
Tr × S^r	3	4.26	1.17
Ss/Tr × S^r	88	3.62	
Between *Ss*	95		
Button (B)	1	1.26	1.26
Time (T)	4	16.91	52.84**
B × T	4	.10	. . . .
B × Tr	3	.78	.78
B × S^r	1	23.84	23.84**
T × Tr	12	.26	. . . .
T × S^r	4	.19	. . . .
B × S^r × Tr	3	.69	.69
T × S^r × Tr	12	.24	. . . .
B × T × Tr	12	.26	1.52
B × T × S^r	4	4.83	28.41**
B × T × S^r × Tr	12	.15	. . . .
Ss × B/Tr × S^r	88	1.00	
Ss × T/Tr × S^r	352	.32	
Ss × B × T/Tr × S^r	352	.17	
Within *Ss*	864		
Total	959		

** $p \leq .01$.

significant differences in total number of extinction responses for those groups given tokens in extinction and those who received nothing. All groups extinguished, i.e., there was a significant trend over time to fewer and fewer responses ($p < .01$). In addition, the Button × Secondary Reinforcement and the Button × Secondary Reinforcement × Time interactions were significant ($p < .01$), as may be seen in Figure 1. This figure shows the mean number of responses made on the token button and on the candy button during each minute of extinction for the four groups which received S^r in training and extinction (S^r-S^r) and for the four groups which received S^r during training trials, but none in extinction (S^r-O). The solid lines connect the mean number of responses to the *candy* button for the S^r-S^r (open circles) and S^r-O (closed circles) groups separately. The S^r-S^r groups made almost four times as many candy-button presses in the first minute of extinction as did the S^r-O groups, but made about the same number of

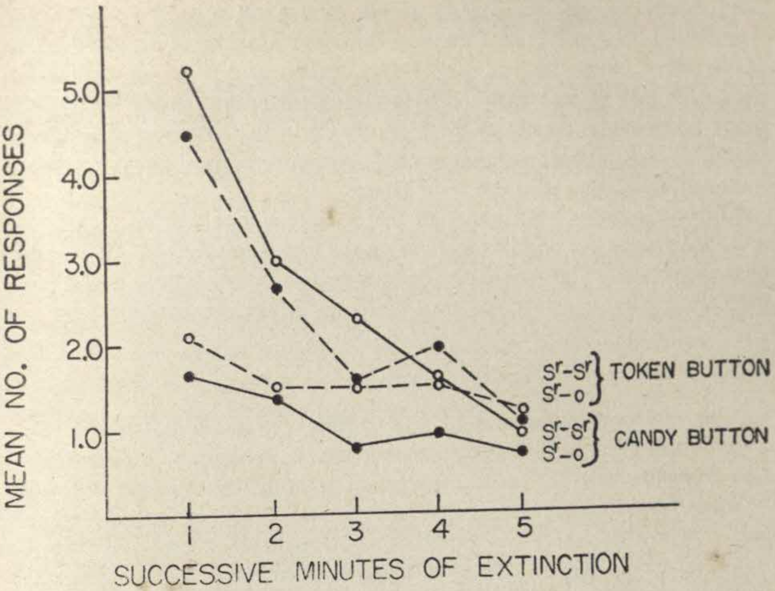

FIGURE 1—Mean number of responses on each button for each minute of extinction for the four E-groups receiving tokens in extinction (S^r–S^r) and for the four E-groups receiving nothing in extinction (S^r–0).

TABLE 2

NUMBER OF RESPONSES ON EACH BUTTON FOR EACH MINUTE OF EXTINCTION FOR S^r-S^r, S^r-O, AND P-O GROUPS

	MINUTES OF EXTINCTION					
	1	*2*	*3*	*4*	*5*	*Total*
S^r-S^r *Groups*						
On Token Button	101	72	73	74	58	378
On Candy Button	251	145	111	77	45	629
Total	352	217	184	151	103	1007
S^r-O *Groups*						
On Token Button	215	129	75	96	53	568
On Candy Button	82	65	38	45	31	261
Total	297	194	113	141	84	829
P-O *Groups*						
On Token Button	9	10	22	18	3	62
On Candy Button	293	208	209	174	161	1045
Total	302	218	231	192	164	1107

candy-button presses as the S^r-O groups in the last minute of extinction. The dotted lines in Figure 1 connect the mean number of responses made to the *token* button for the S^r-S^r (open circles) and S^r-O (closed circles) groups. It may be seen that the S^r-O groups made approximately twice as many token-button presses as the S^r-S^r groups in the first minute of extinction. The S^r-S^r groups made more candy-button responses and fewer token-button responses than did the S^r-O groups.

Table 2 presents a summary of the extinction data for the S^r-S^r, S^r-O, and P-O groups. An analysis comparing the total number of button presses of the experimental groups with the four groups (P-O) who received no S^r even during training revealed no differences in performance.

DISCUSSION

The variation in number of trials before extinction did not produce a significant effect in the present experiment. A prediction of more extinction responses with increased numbers of training trials might ordinarily be made for a learning situation. The response in this study, however, was one well established in the repertoire of preschoolers, and the situation might be more accurately termed a performance one. Lewis and Duncan (5) first commented on this distinction and predicted, for an overlearned response, that increased training trials would decrease the number of extinction responses, due to a facilitation of discrimination between training and extinction. The number of training trials in this instance varied between two and 16; it is possible that 16 regularly reinforced trials were not more effective than two in enabling the *S*s to discriminate extinction from training. All groups showed a very low level of responses in extinction.

Although the groups given tokens in extinction (S^r-S^r) made *more* responses in toto than did the groups extinguished with nothing (S^r-O), the difference merely approached significance in this study. However, it must be remembered that token reinforcement when administered actually produced a delay in button-responding, as it necessitated the operations of retrieving and inserting the token before another button-press was called for. Despite this enforced delay, the token reinforced groups made more responses.

The use of two buttons in this situation demonstrated the "chaining" function of the S^r. The significant Button × S^r interaction revealed that the groups given S^r in extinction made more candy-button presses and fewer token-button presses than did the S^r-O groups. The S^r-O groups received nothing for responding on the token button. This operated to break the established chain of behavior: The *S*s tended, continually and immediately, to repress the token button, which led to a high initial rate of responding on the token button and to more token-button presses for the S^r-O groups than for the S^r-S^r groups. The number of token-button presses per unit time decreased rapidly for the S^r-O groups, however, as seen in Figure 1.

The S^r, when administered in extinction, seemed to release the additional behavior in the chain, i.e., insertion of the token and subsequent pressing of the candy button. Thus, the S^r-S^r groups made more candy-button presses than did the S^r-O groups. Zimmerman (10) has suggested that this release of new behavior patterns is a necessary ingredient for strong S^r effects. In this particular experimental situation, at least, it was not a sufficient condition. However, the S^r seems to have had some negligible effect upon token-button performance. The rather constant rate of S^r-S^r groups' token-button pressing throughout the five-minute extinction period should be noted in Figure 1 and contrasted with the decreasing rate of token-button pressing revealed by the S^r-O groups. If the extinction period had been extended, this suggested difference in trend might have shown more distinctly.

SUMMARY

This experiment was concerned with the extinction behavior of children following varying numbers of reinforced trials. Children were presented individually with a "clown game." When a button on the nose of the clown was pressed, a token was delivered from the clown's mouth. The token was inserted in the slot provided on the clown's face, and a second button on the nose pressed. This time a candy fell from the mouth. There were 12 groups of eight *S*s each, differing in reinforcement and extinction procedures and equated for mean age and sex. Each child received either 2, 4, 8, or 16 training trials with either token *and* candy reinforcement, or candy reinforcement alone, followed by five minutes of extinction. There were no differences in total group performance due to the number of training trials or secondary reinforcement. More candy-button responses were made by groups extinguished under S^r; more token-button responses were made by the groups receiving nothing in extinction.

REFERENCES

1. BERSH, P. J. The influence of two variables upon the establishment of a secondary reinforcer for operant responses. *J. exp. Psychol.*, 1951, 41, 62-73.
2. CAPALDI, E. J. Effect of different amounts of alternating partial reinforcement on resistance to extinction. *Amer. J. Psychol.*, 1957, 70, 451-452.
3. CAPALDI, E. J. Effect of different amounts of training on resistance to extinction of different patterns of partially reinforced responses. *J. comp. physiol. Psychol.*, 1958, 51, 367-371.

4. Hall, J. F. Studies in secondary reinforcement: I. Secondary reinforcement as a function of the frequency of primary reinforcement. *J. comp. physiol. Psychol.*, 1951, 44, 246-251.

5. Lewis, D. J., & Duncan, C. D. Expectation and resistance to extinction of a lever-pulling response as a function of percent reinforcement and number of acquisition trials. *J. exp. Psychol.*, 1958, 55, 121-128.

6. Miles, R. C. Relative effectiveness of secondary reinforcement throughout deprivation and habit strength parameters. *J. comp. physiol. Psychol.*, 1956, 49, 126-130.

7. Myers, N. A. Extinction following partial and continuous primary and secondary reinforcement. *J. exp. Psychol.*, 1960, 60, 172-179.

8. Siegel, P. S., & Foshee, J. G. The law of primary reinforcement in children. *J. exp. Psychol.*, 1953, 45, 12-14.

9. Spiker, C. C. The effects of number of reinforcements on the strength of a generalized instrumental response. *Child Develpm.*, 1956, 27, 37-44.

10. Zimmerman, D. W. Durable secondary reinforcement: method and theory. *Psychol. Rev.*, 1957, 64, 373-383.

11. Zimmerman, D. W. Sustained performance in rats based on secondary reinforcement. *J. comp. physiol. Psychol.*, 1959, 52, 353-358.

EFFECTS OF ASSOCIATIVE STRENGTH, GRADE LEVEL, AND INTERPAIR INTERVAL IN VERBAL PAIRED-ASSOCIATE LEARNING[1]

JOHN C. MCCULLERS*[2]

University of Texas

Recent work by Castaneda, Fahel, and Odom (1) has led to a list of words for investigating verbal learning phenomena among elementary school age children by means of the paired-associate method. Although these investigators report associative characteristics for a population of fourth-, fifth-, and sixth-grade children, they do not indicate what, if any, differences in association values arise as a function of grade level. The associative strengths between the word pairs reported by the authors might reasonably be expected to be a function of the number of times the child, during his past experience, has formed an association between any two given words. Accordingly, different associative strengths as a function of grade level might be predicted on the basis of the more general factor of years of linguistic experience as well as the more specific factor of formal (classroom) training in reading and the basic language skills. Thus, if associative strength between the word pairs varies as a function of grade level, differences in paired-associate learning would also be expected.

Furthermore, it may be noted that the original study employed only a single temporal interval (2.0 sec.) between the presentation of word pairs. The present study varies the interpair interval at four levels as well as the factors of grade level and association value and should provide infor-

---

* Institute of Child Development, University of Minnesota, Minneapolis 14.

[1] This study was supported in part by a Public Health Service Predoctoral Research Fellowship (MR-13000) from the National Institute of Mental Health to the author and in part by grants from the Hogg Foundation, the University of Texas, and the National Institute of Mental Health (Grant M-4240) to Dr. Alfred Castaneda, under whose supervision the present research was carried out.

[2] The author wishes to express his appreciation to the directors of the Austin Public Schools, to Mr. Fred Matthys, Principal, Highland Park Elementary School, Austin, Texas, and to Mrs. Leila Fahel of the Office of Child Research, University of Texas, for their assistance and cooperation.

Child Develpm., 1961, 32, 773-778.

mation regarding either the main or the interactive effects among these three variables. The present study, then, may be viewed as an extension of the Castaneda, Fahel, and Odom investigation with a twofold purpose of providing information relevant to the questions raised above and also of serving as a further check on the usefulness of the word association list described in the initial study.

METHOD

Subjects

The subjects consisted of 80 elementary school children selected on the basis of grade level; two groups of 40 subjects each were selected at the fourth and sixth grade levels.

Apparatus and Materials

The apparatus used for exposing the words was the Hunter Card Master (Model 340). A 3-sec. anticipation interval during which the stimulus word was exposed preceded a 3-sec. exposure of the response word. There was, depending on the particular experimental condition, either a 2-, 3-, 4-, or 6-sec. interval between the exposures of the pairs of words. The words used to form the experimental lists for both the high and low association conditions were chosen so as to minimize response competition within the lists. This was accomplished by placing the following restrictions on the selection of words from the normative list: no word was selected for use which (a) carried associative strength, as determined from responses in the normative study, for any word in the list other than its associate or which (b) carried associative strength mediated through some neutral word for any other word in the list except its associate. The final lists were, as nearly as possible, made alike in all respects except in the associative strengths of the word pairs. The lists of the word pairs and their association values are given in Table 1.

TABLE 1

WORD PAIRS AND THEIR ASSOCIATION VALUES (in parentheses)

	RESPONSE	
Stimulus	*Group HI*	*Group LO*
LITTLE	SMALL (60)	THIN (2)
ANGRY	MAD (61)	CRUEL (2)
SLOW	FAST (50)	RUN (5)
AFRAID	SCARED (61)	KNIFE ((2)
WARM	HOT (57)	SUMMER (2)
BEAUTIFUL	PRETTY (62)	HOUSE (2)

Procedure and Design

Each *S* was brought individually into the experimental room and seated in front of the apparatus. The apparatus was located on a table at a height which facilitated *S*'s viewing of the exposed words.

As a precautionary measure to ensure that *S*'s rate of learning would not be hindered by difficulties in reading and understanding, *E* began by asking *S* to read aloud a list of 36 words which were prepared in alphabetical order and which included the 12 words that were to be used in the task. The 24 words on the list which were not used in the task were selected from the normative group of words in such a way that none of these words had association values indicated for any of the 12 words used in the task. The *E* corrected pronunciation and asked *S* to repeat the list in case of one or more reading errors.

Since the instructions vary slightly from those of the original study, they are presented verbatim as follows:

> Now I will read the instructions to you. This is a game to see how well you can remember. This is how it goes. Here on this side (*E* points to the left aperture) a little window will open up, and you will see a word. A few seconds after this window opens, the window on this side (*E* points to right aperture) will open and you will see a *second* word. Your job is to guess what the *second* word is before its window opens. You say the word out loud so that I can hear you. Some of the words on the list that you read will be used in the machine. I had you read the list so that I would know that you could read anything that turns up in the machine. Do you think you understand how it goes now? Fine; now we are ready to start. Remember, you are to say out loud the second word which is going to appear in the second window *before* the window opens. In the beginning you will make a lot of mistakes, but as you learn how it goes, you won't make so many mistakes.

The subjects at each grade level were randomly assigned either to a group which was presented six pairs of words all having high association value (Group HI) or to a group which was presented six word pairs having low association value (Group LO).

Four conditions differing in the length of the interpair interval were used. The interpair interval was considered to be that interval which occurred between the exposures of separate pairs of words and in which both windows of the apparatus were closed. Interpair intervals of 2, 3, 4, and 6 sec. were employed. A 2×2×4 factorial design consisting of 16 cells or conditions was thus generated by the variables of grade level, association value of the words, and the length of the interpair interval. Five *S*s were assigned to each of the eight conditions within each grade level. The groups were fairly evenly balanced with respect to sex.

After giving the instructions to *S*, *E* seated himself to the side of *S* in such a way as to be able to see the windows of the apparatus and did not further interact with the *S*. Each child was carried to a criterion of one errorless trial or to a maximum of 15 trials. A trial consisted of one exposure of each of the six word pairs. In order to prevent the word pairs from being learned serially, three different orders of exposing the pairs were employed.

Results and Discussion

The mean number of trials to criterion was computed for the *S*s in each condition defined by grade level, association value, and interpair interval. This information, which is presented graphically in Figure 1, reveals a tendency toward a grade-level effect. As may be seen, with the exception of the low association condition at the 6-sec. interval, sixth graders reached criterion earlier than fourth graders at every level of interpair interval for both conditions of association. Inasmuch as all *S*s reached the criterion of one errorless trial, a 2 × 2 × 4 factorial analysis of variance was performed using the number of trials to criterion as each individual's score.[3] The results of this analysis indicated that the factor of grade level did not achieve statistical significance ($F = 3.15$, $df = 1, 64$). The *S*s for this study were selected at the fourth and sixth grade levels, which constituted the extremes in grade level of the population studied by Castaneda, Fahel, and Odom, on the assumption that any grade-level effect which might be present would be uncovered. Odom (2) also failed to find a significant effect of grade level in a study employing word pairs from the same normative list. The results of these two studies, when considered together, tend to support the notion that for paired-associate tasks among fourth, fifth, and sixth graders, the word association values reported by Castaneda, Fahel, and Odom are not significantly influenced by grade level.

Figure 1 also shows the sharp differences obtained in rate of learning as a function of the association value of the words used. This difference between the HI and LO groups proved to be significant well beyond the .001 level ($F = 26.31$, $df = 1, 64$). This finding furnishes additional experimental verification of the general utility of the original word association list as a means of varying the parameter of associative strength.

Interpair interval proved to be a nonsignificant factor with the particular intervals chosen. It should be pointed out, however, that while the interpair interval normally employed in adult studies is of a duration of less than 1 sec., the shortest interval used in the present study, due to physical limitations of the apparatus, was 2 sec. Thus, all of the interpair intervals employed in the present study were equal to or longer than the one employed by Castaneda, Fahel, and Odom. Therefore, while it may be concluded that increasing the interpair interval by as much as 4 sec. does not appear to affect significantly rate of learning, the possibility remains that shortening the length of this interval below 2 sec. may produce a significant effect. In line with this, interpair interval might also be expected to interact with anticipation interval and perhaps the duration of time in which the response item is exposed.

---

[3] The same analysis performed using errors as scores yielded essentially the same results.

Figure 1 further reveals a slight tendency toward a triple interaction effect. There is also a tendency, particularly in the fourth graders, for poorer performance to be associated with increasing length of the interpair interval. The statistical analysis, however, indicated that none of the interactive effects approached significance.

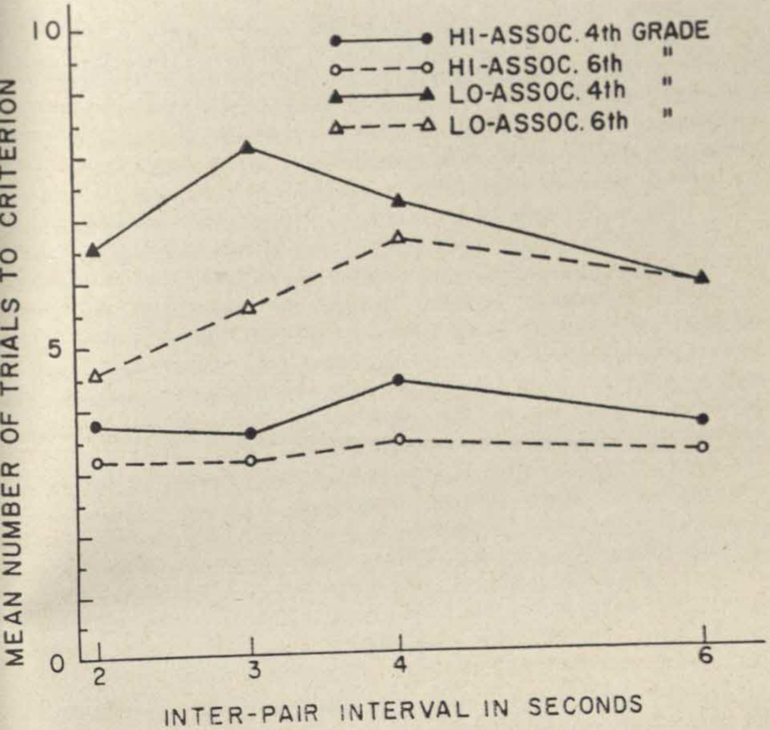

FIGURE 1—Mean number of trials to criterion plotted separately for each condition.

SUMMARY

The present study investigated verbal paired-associate learning among 80 school children at the fourth and sixth grade levels by means of the anticipation method. The variables examined were grade level, association value of the word pairs, and interpair interval. Two levels of association value were employed, a high and a low association condition, and the interpair interval was studied at four levels. The only factor reaching statistical significance proved to be the main effects factor of association value. Some

discussion and evidence are offered concerning the negative findings with regard to grade level.

REFERENCES

1. CASTANEDA, A., FAHEL, L. S., & ODOM, R. Associative characteristics of sixty-three adjectives and their relation to verbal paired-associate learning in children. *Child Develpm.*, 1961, 32, 297-304.

2. ODOM, R. Relation of association-value, competition and grade level to verbal paired-associate learning in children. Unpublished master's thesis, Univer. of Texas, 1960.

BEHAVIOR OF YOUNG CHILDREN IN A TWO-CHOICE GUESSING PROBLEM [1,2]

WILLIAM KESSEN* *and* MARION LORD KESSEN
Yale University

Humphreys' basic research (6) on conditions of uncertain outcome has been followed up along several interrelated lines—partial reinforcement studies (*see* 7), tests of mathematical models of probability learning (*see* 3), and studies of cognitive structure (e.g., 4, 5). Several recent studies (8, 10, 11) have extended the range of these settings to include the behavior of children. Convincing regularities have been demonstrated in the asymptotic behavior of animals and men in situations of uncertain outcome, and recent research has turned to an investigation of the variables other than degree of uncertainty which control asymptotic performance (e.g., 9) and to the study of transitional probabilities in response patterns (e.g., 12).

The present study developed from an interest in the following questions: (a) Do young children show the regularities in performance which have been shown in adults? (b) Does the behavior of children in an uncertain-outcome situation vary with prior experience in such situations? (c) Does prior experience have different effects at different ages? (d) What are the patterns of two-choice guessing in young children; i.e., what do the "rules of solution" or "strategies" look like when expressed in terms of transitional probabilities of responding? The research was designed to point up developmental variation and its implications for theories of behavior in situations of uncertain outcome; no formal mathematical model for probability learning, dealing either with asymptotic performance or with sequential dependencies, was under test.

METHOD

Subjects

Seventy-two children from two age-graded classes in each of two nursery schools served as *Ss*.[3] Within each age group (called group A and group B below), children were assigned at random to one of four experimental

---

* Department of Psychology and Child Study Center, Yale University, New Haven, Connecticut.

[1] This research was supported in part by USPHS Research Grant M-1787 and was completed while the first author was a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, California.

[2] The authors want to thank Richard C. Atkinson, John Gilbert, and George Mandler, who helped in the preparation of this report in several unduplicable ways.

[3] Miss Omwake and her staff at the Yale University Child Study Center Nursery School and Mrs. Norman and her staff at the Yale University Divinity School Nursery School, both of New Haven, Connecticut, gave freely of their time and convenience in providing *Ss* for the research.

Child Develpm., 1961, 32, 779-788.

groups (called 50/50, 50/70, 70/50, and 70/70 below). Because no significant variation was found between the children from different schools, their data are reported together. Table 1 shows the distribution of *S*s by age group, school, and sex. The median age of *S*s in group A was 3 years, 7 months; the median age of *S*s in group B was 4 years, 5 months. Furthermore, accepting the nursery schools' division into classes left only three *S*s of group A over 4 years old and only five *S*s of group B under 4 years old.

TABLE 1

DISTRIBUTION OF SUBJECTS BY AGE GROUP, SEX, AND SCHOOL

	GROUP A Median Age 3-7		GROUP B Median Age 4-5	
	Boys	*Girls*	*Boys*	*Girls*
School 1	9	7	8	11
School 2	9	11	7	10

Materials

Blank-faced playing cards were printed with one of two designs—a solid black rectangle 1 inch by 2 inches or a solid red rectangle of the same dimensions. The cards were arranged in three packs of 50 cards each; two packs had an over-all distribution of 25 cards with black rectangles and 25 cards with red rectangles, and the third pack had 35 cards with black rectangles and 15 cards with red rectangles. A further restriction was imposed: for each block of 10 cards, the over-all distributions were the same (i.e., either five black and five red cards or seven black and three red cards), a restriction which limits the length of runs of one design compared to packs drawn without such restrictions. The sequence of cards with black (B) and red (R) rectangles in the 35-15 pack was BRBRBBRBBB BBBRBRBBBR BRBBBBRBRB RBRBBBBRBB BBRRBBRBBB. The sequences of the two packs of 25 B and 25 R were RRBBRBBBRR RBRBRBBBRR RBRBBBRBRR RBRRRBBBRB BBRRRBBBRR and BRBBBRBRRR BRBRRBRBBR RBBRRBBBRR BBRRBRBBRR BRRBBRRBBR.

Procedure

Each *S* was taken by *E* from his nursery school classroom to another room in the same building and presented with the task individually. With *E* seated on one side of a table and *S* seated on the other, *S* was told: "We are going to play a game. I have some cards here. (*E* exposed two cards illustrating the red and black rectangle). The cards are like these. All of the cards are either red or black. Do you know what color this is? (*E* indicated one of the cards). And what color is this? (*E* indicates the other card). Now what I want you to do is tell me what color you think the next card in the pack will be. What color do you think the next card is going to be?" These instructions were varied slightly to suit the child, but invariably contained the notions of game, of predicting the color of the next card, and an inquiry to determine the child's color recognition. *S* then said either "Red" or "Black," and the first card was turned so that he could see it. He then said either "Red" or "Black," the next card was exposed, and so on through 50 cards. After each card had been shown to *S*, *E* put it on top of the preceding card. After the first 50 cards had been shown, *E* took a pack of cards and said to *S*: "Now we have some more cards. What color do you think this one will be?" The second pack of 50 cards was then run through in the same way as was the first.

Most children responded well to the task and cooperated with *E*, though with some reluctance on occasion. For the children in this study and with the materials used, 100 guesses seemed close to *Ss*' limits of interest and attention.

The four experimental groups varied in the composition of the packs of cards to which they were exposed. Group 50/50 predicted and saw exposed a pack made up of 25 cards with black rectangles and 25 cards with red rectangles on both the first and second runs of 50 cards. Group 50/70 predicted and saw exposed a pack of 25 B and 25 R on the first run and a pack of 35 B and 15 R on the second run. Group 70/50 was exposed to a pack of 35 B and 15 R on the first run and a pack of 25 B and 25 R on the second run. Group 70/70 was exposed to a pack of 35 B and 15 R on both the first and second runs. Thus, for each age group, one experimental group had 100 exposures of a half black - half red composition, one had 100 exposures of a 70 per cent black - 30 per cent red composition; one had 50 exposures of half black - half red composition and was shifted to 50 exposures of 70 per cent B and 30 per cent R; and finally, one experimental group at each age level first had 50 exposures of 70 per cent B and 30 per cent R followed by 50 exposures of half black - half red composition.

RESULTS

Level of Performance

The performance of groups A and B over 10 blocks of 10 responses each is shown in Figure 1 and Figure 2. The performance of groups 50/50

TABLE 2

ANALYSIS OF VARIANCE OF BLACK RESPONSE, TRIALS 41–50

Source	*df*	*MS*	*F*	*p*
Pack composition (P-c)	1	25.68	4.93	$< .05$
Age (A)	1	2.35	< 1	
P-c × A	1	.35	< 1	
Within groups	68	5.20		

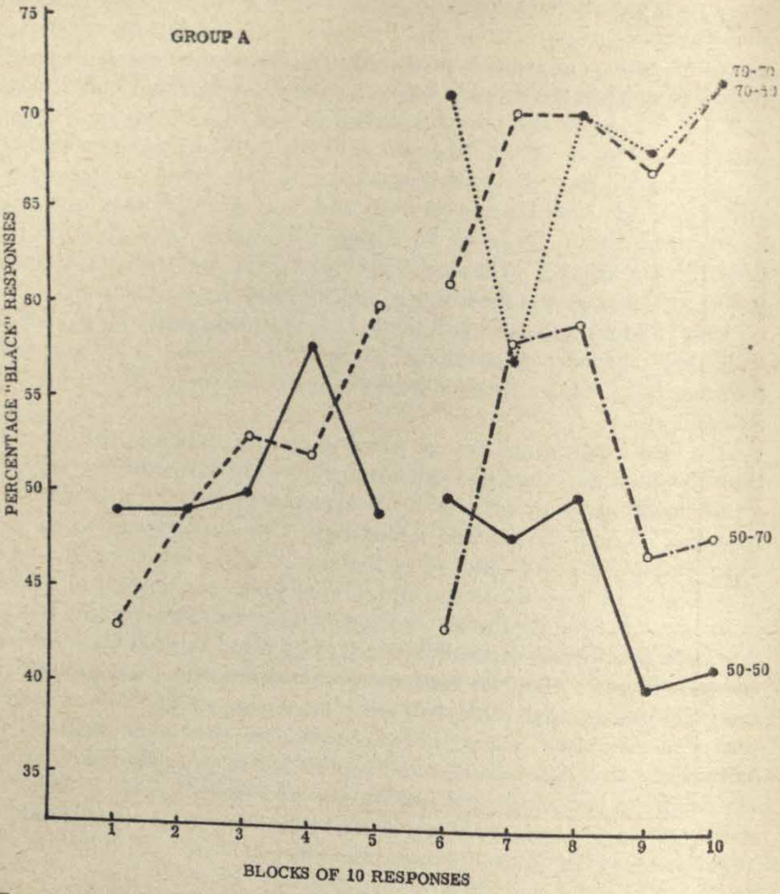

FIGURE 1—Percentage of "Black" responses for younger children (group A). Dashed curve for first five blocks shows responses to 35-15 pack; solid line shows responses to 25-25 pack.

and 50/70 is shown in a single line for the first 50 responses, as is the performance of groups 70/50 and 70/70. At the end of the first fifty guesses and card exposures, both age groups appear to be responding differentially to the 50 per cent and 70 per cent sequences of cards, with the absolute difference in percentage of "Black" response being larger for the older than for the younger children. Table 2 presents an analysis of variance of performance on trials 41 to 50 and indicates that the composition of the packs made a significant difference in performance, but that neither age nor the interaction of age with pack composition approaches statistical significance.

After 100 guesses and card exposures, the groups that had two runs of the same pack (group 50/50 and group 70/70) show different perform-

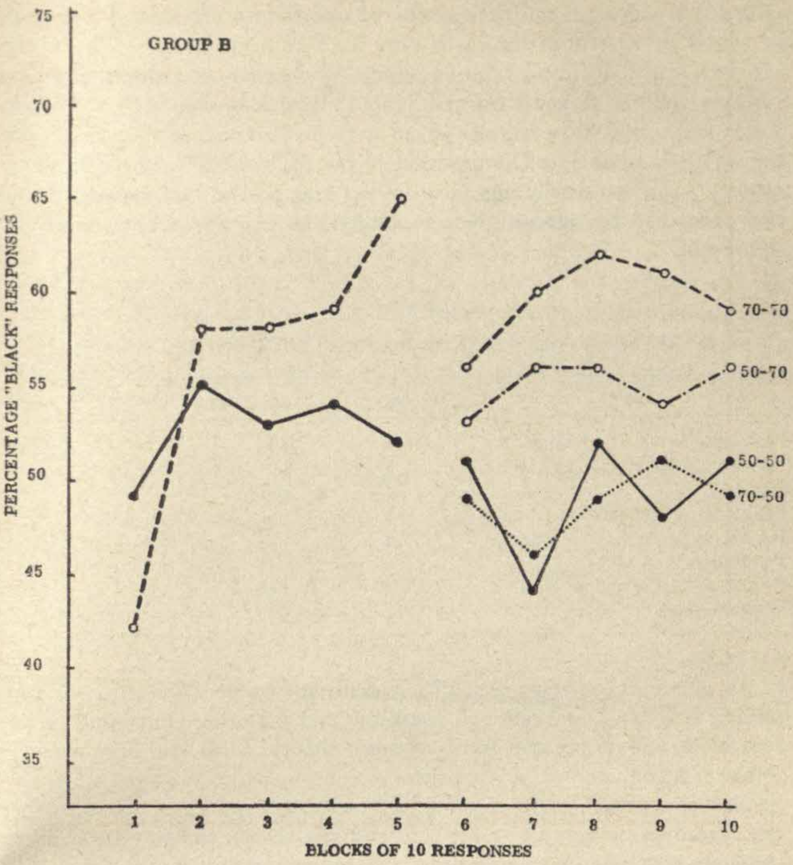

FIGURE 2—Percentage of "Black" responses for older children (group B). Dashed curve for first five blocks shows responses to 35-15 pack; solid line shows responses to 25-25 pack.

ance levels ($p < .01$). It is the behavior of the 50/70 and the 70/50 groups that is of particular interest, however, and particularly among the younger children. As can be seen from Figure 2, the children in group B who were shifted from one pack to another responded to the shift by changing their guessing behavior to conform with the composition of the *second* pack (i.e., they responded "Black" more often to the 35-15 pack than to the 25-25 pack), although the difference between group 70/50 and group 50/70 is less than that between group 70/70 and group 50/50, presumably as a consequence of varying prior experience. The children in group A, as can be seen in Figure 1, behaved in a paradoxical way to the shift in packs. Both the 70/50 and the 50/70 groups performed in a way appropriate to the *first* pack; thus, the children who guessed and saw exposed a pack of cards half with black rectangles and half with red and then were shifted to a pack in which 70 per cent of the cards were black continue to guess "as if" the cards were still half and half; a similar persistence was shown by these younger children shifted from the 35 B - 15 R pack to the 25 B - 25 R pack. Table 3 describes these results in an analysis of variance of performance during trials 91 to 100. Composition of the pack used for the first 50 responses has a statistically significant effect and, as the figures suggest, the *interaction* between age and the composition of the first pack was statistically significant.

TABLE 3

ANALYSIS OF VARIANCE OF BLACK RESPONSE, TRIALS 91–100

Source	*df*	*MS*	*F*	*p*
Pack composition (P-c) 1st 50	1	36.13	6.71	< .05
Pack composition (P-c) 2nd 50	1	5.01	< 1	
Age (A)	1	4.01	< 1	
P-c 1st 50 × P-c 2nd 50	1	.01	< 1	
P-c 1st 50 × A	1	33.35	6.20	< .05
P-c 2nd 50 × A	1	.69	< 1	
P-c 1st 50 × P-c 2nd 50 × A	1	3.06	< 1	
Within groups	64	5.38		

In summary, *S*s responded differentially to two sets of cards, one containing half black and half red markings and the other containing 70 per cent black and 30 per cent red markings, after 50 cards had been exposed. When a second run of 50 exposures was made with the same packs, this result was repeated. In the case of children shifted from one pack to another (i.e., from 50 per cent to 70 per cent and vice versa), the younger children showed a strong tendency to respond in a way appropriate to the distribution of the first-seen pack. Stated in another way, the younger children are less willing to give up an operating rule or response tendency once it has been established.

Transitional Probabilities

Developmental variation of a somewhat different order can be seen in an examination of transitional probabilities of responding "Black" or "Red." How do the patterns or "rules" of response sequence vary with age or prior experience? Suppes and Atkinson (12) have presented an adaptation of a technique of analysis introduced by Anderson and Goodman (2) which is appropriate to this question. Each response of each child (i.e., on every trial n), whether "B" or "R," was tabulated according to whether on the preceding trial (trial $n-1$) his response had been "B" and correct (B+), "R" and correct (R+), "B" and incorrect (B−), or "R" and incorrect (R−). From this basis, the likelihood of saying "B" or "R" after each of these four prior events was computed. Table 4 shows the resulting tabulation of transitional probabilities for the last 40 responses of each run for both age groups and both pack compositions. Thus, Table 4 presents a summary statement of *S*s' tendency to respond "B" or "R" (a) as this tendency relates to how they responded on the immediately preceding trial and (b) as it relates to whether or not *S*s had been correct on their immediately preceding guess. For example, the figures in the upper lefthand corner of Table 4 show that, for the younger group of children seeing the 25-25 pack for the first time, the response "B" occurred 56 per cent of the time after the response "B" had occurred on the preceding trial and was correct. The same tabulation shows that the response "R" occurred only 39 per cent of the time after "R" had occurred on the preceding trial and been wrong. The behavior of the younger and older groups can be compared by examining the probability values shown in the top half of the table (group A) with those shown in the bottom half (group B).

Table 4

PATTERNS OF RESPONSE SEQUENCE (TRANSITIONAL PROBABILITIES) FOR ALL GROUPS IN ALL CONDITIONS

(Trials 11 to 50 and 61 to 100 are represented.)

					Second Run							
	First Run				25-25 Pack				35-15 Pack			
On Trial $n-1$	25-25 Pack		35-15 Pack		*After* 25-25		*After* 35-15		*After* 25-25		*After* 35-15	
	On Trial *n*		On Trial *n*		On Trial *n*		On Trial *n*		On Trial *n*		On Trial *n*	
	B	R	B	R	B	R	B	R	B	R	B	R
Group A												
B+ . .	.56	.44	.60	.40	.43	.57	.82	.18	.54	.46	.73	.27
R+ . .	.53	.47	.65	.35	.51	.49	.39	.61	.69	.31	.97	.03
B− . .	.59	.41	.68	.32	.53	.47	.84	.16	.79	.21	.65	.35
R− . .	.39	.61	.32	.68	.33	.67	.29	.71	.31	.69	.54	.46
Group B												
B+ . .	.35	.65	.51	.49	.34	.66	.29	.71	.46	.54	.49	.51
R+ . .	.74	.26	.86	.14	.68	.32	.60	.40	.80	.20	.91	.09
B− . .	.40	.60	.45	.55	.24	.76	.26	.74	.34	.66	.62	.38
R− . .	.73	.27	.71	.29	.70	.30	.81	.19	.66	.34	.64	.36

These comparisons were subjected to the statistical analysis suggested by Suppes and Atkinson, with the outcomes shown in Table 5. In all six over-all comparisons of the transitional behavior of the two age groups (e.g., first run of 25-25 pack for group A against first run of 25-25 pack for group B), significant differences appeared. Table 5 also shows the significance of the differences in behavior of the two age groups for *each preceding event* (i.e., B+, B—, R+, R—). Thus, the difference between the behavior of group A and that of group B after responding "R" on the preceding trial and being wrong (row R—) is highly significant in five of the six situations.

For the table as a whole 15 of the 24 comparisons between age groups are significant beyond the .01 level of chance expectation, three are beyond the .05 level, and six fall short of statistical significance.

TABLE 5

CHI SQUARE VALUES ($df = 1$) FOR DIFFERENCES IN TRANSITIONAL PROBABILITIES OF RESPONSE BETWEEN GROUP A AND GROUP B

	FIRST RUN		SECOND RUN			
	25-25	35-15	25-25 *after* 25-25	25-25 *after* 35-15	35-15 *after* 25-25	35-15 *after* 35-15
B+	16.82**	4.75*	1.58	59.61**	1.62	20.88**
R+	15.73**	11.12**	5.52*	6.54*	1.63	1.08
B−	14.55**	11.65**	14.71**	71.59**	22.44**	.21
R−	40.29**	63.04**	26.49**	40.55**	28.71**	1.62

* $p < .05$.
** $p < .01$.

Examination of Table 4 suggests several summary conclusions. First, the older children were more stable in the application of their operating principles to packs of different composition. For group B, the three sets of transitional probabilities for the 25-25 pack show marked similarity, as do the three sets of transitional probabilities for the 35-15 pack. This regularity is not readily apparent in group A. Second, the older children alternated their responses more frequently than did the younger children. This difference can be seen throughout Table 4 by comparing the likelihood that "B" will follow "R" and that "R" will follow "B" with the likelihood that two "B's" or two "R's" will occur consecutively. In 23 of the 24 possible comparisons shown in Table 4, the younger children follow "B" with "B" and "R" with "R" at higher probability than do the older children and follow "B" with "R" and "R" with "B" at lower probability. Furthermore, if it is recalled that 18 of these comparisons are statistically significant (the single reversal is not), the age difference in the tendency to alternate would seem to be established. Third, the relative tendency among the older children to alternate is particularly in evidence when they have made an incorrect prediction. Compare, for example, the response of group A to the first run of the 25-25 pack when incorrect on the preceding trial with the corresponding pattern of group B. After saying "B" and being wrong, the younger children shifted their response to "R" 41 per cent and repeated "B" 59 per cent while the older children shifted their response to "R" 60 per cent and repeated "B" 40 per cent; correspondingly, after saying "R" and being wrong, younger children shifted 39 per cent and repeated 61 per cent while older children shifted 73 per cent and repeated 27 per cent. This tendency for the younger children to repeat when wrong and for the older children to shift response when wrong runs through the data. To assess this differential tendency, all *S*s were scored over trials 11 to 50 as to whether they alternated more frequently when correct on the immediately preceding trial or when incorrect on the immediately preceding trial. Twenty-three of the 36 younger children alternated more often when correct; only 14 of the 36 older children did so. This difference is statistically significant at the .05 level of confidence. As was the case in the analysis of level of performance, the younger children show reluctance in giving up the repetition of a guess even in the face of disconfirming evidence.

In summary, the differences between the two age groups in their "rules of guessing" are large, statistically significant, and regular in regard to the tendency to alternate or to repeat when wrong on the preceding guess.[4] From the data in hand, it cannot be determined whether these differences represent a systematic "development" toward adult patterns of behavior in the two-choice setting. Only the analysis of transitional response probabilities over a wider age range will permit statements about general trends in the effects of age and its correlates on guessing behavior.

These results furnish a demonstration of differences in behavior in the two-choice situation which cannot be dealt with solely by a consideration of the stimulus materials to which the *S*s were exposed. The differences between age groups point up the interactive determination of behavior in situations of uncertain outcome. Not only is variation in behavior related to the materials used, but variation is also related to the patterns of guessing that *S*s bring to the task.[5] Under the conditions used in the present study, there appears to be a dramatic shift in the strategy of predicting uncertain outcomes near the end of the fourth year of life.

SUMMARY

Seventy-two preschool children in two age groups predicted the sequential occurrence of cards marked with either a black (B) or a red (R) rectangle. The behavior of four experimental groups, in a two-by-two design for each age group, was studied; one group predicted a pack of 50 cards that was 50 per cent B - 50 per cent R and then predicted a pack of 50 cards 70 per cent B - 30 per cent R, one group predicted the packs in the reverse order, and two groups predicted either 50 per cent B - 50 per cent R or 70 per cent B - 30 per cent R packs in both runs of 50 cards each.

The children responded differentially to the 50 per cent and 70 per cent packs. The older children who were shifted from the 50 per cent pack to the

---

[4] The examination of the influence of events more remote than the preceding trial can be properly conducted only with longer series of responses than could be obtained in the present study.

[5] Anderson (1), in a study of the effect of first-order conditional probabilities, has presented a related argument for the closer study of individual differences in patterns of alternation and repetition.

70 per cent pack or from the 70 per cent pack to the 50 per cent pack showed a tendency to shift their guessing behavior appropriately; the younger children who were shifted from one pack to another showed a paradoxical tendency to continue guessing in a way appropriate to the *first* run of cards.

An analysis of transitional probabilities of responding showed clear differences between the two age groups. The younger children were more likely to repeat the response made on the preceding trial than were the older children, and this difference was more clearly seen when the guess made on the preceding trial was incorrect.

The results were interpreted as demonstrating the interaction of stimulus materials and strategies of prediction in the determination of behavior in situations of uncertain outcome.

REFERENCES

1. ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. *J. exp. Psychol.,* 1960, 59, 73-93.
2. ANDERSON, T. W., & GOODMAN, L. A. Statistical inference about Markov chains. *Ann. math. Statist.,* 1957, 28, 89-110.
3. ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: a study of a science.* Vol. 2. McGraw-Hill, 1959. Pp. 380-491.
4. FELDMAN, J. An analysis of predictive behavior in a two-choice situation. Unpublished doctoral dissertation, Carnegie Institute of Technology, 1959.
5. GOODNOW, J. J. Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.,* 1955, 68, 106-116.
6. HUMPHREYS, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.,* 1939, 25, 294-301.
7. JENKINS, W. O., & STANLEY, J. C., JR. Partial reinforcement: a review and critique. *Psychol. Bull.,* 1950, 47, 193-234.
8. MESSICK, S. J., & SOLLEY, C. M. Probability learning in children: some exploratory studies. *J. genet. Psychol.,* 1957, 90, 23-32.
9. SIEGEL, S., & GOLDSTEIN, D. A. Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychol.,* 1959, 57, 37-42.
10. STEVENSON, H. W., & WEIR, M. W. Variables affecting children's performance in a probability learning task. *J. exp. Psychol.,* 1959, 57, 403-412.
11. STEVENSON, H. W., & ZIGLER, E. F. Probability learning in children. *J. exp. Psychol.* 1958, 56, 185-192.
12. SUPPES, P., & ATKINSON, R. D. *Markov learning models for multiperson situations.* Stanford Univer. Press, 1960.

EFFECTS OF DELAYED REINFORCEMENT ON ASSOCIATIVE AND INCENTIVE FACTORS[1]

ROBERT WARE *and* GLENN TERRELL*

University of Colorado

Numerous experiments both at the human and infrahuman levels have demonstrated that acquisition is in part a function of the temporal nearness of reinforcement to the response (8, 9, 10, 11). Some effort has been made to determine why immediate reinforcement results in more efficient learning. Perin (9) offered a generalized extinction-inhibition hypothesis to account for the detrimental effect of delayed reinforcement on acquisition. Logan (8) and Lipsitt and Castaneda (7) theorized that the time interval separating reinforcement from reaction affects *J*, which is positively related to the classical conditioned fractional anticipatory goal response, $r_g—s_g$. Logan demonstrated that delayed reinforcement affects *H* in rats. Terrell and Ware (11) suggested that both associative and incentive factors are affected by the time interval separating reinforcement from reaction. In the Terrell and Ware experiment, a discrimination learning study, it was reported on the basis of casual observations that *S*s (a) appeared to take less time to decide which one of several stimulus objects to choose and (b) were apparently more interested in the tasks and pushed the response buttons more quickly when performing under an immediate than a delayed reinforcement condition.

It is believed that, in a discrimination learning study, the time interval between the presentation of a stimulus complex and the *onset* of *S*s' response can reasonably be conceived of as an index of the strength of the

---

* Department of Psychology, University of Colorado, Boulder.

[1] This investigation was supported by a research grant, M-2063, from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service. Grateful acknowledgment is due Nevin J. Platt, Melvin Wiesley, Mrs. Earl W. Almgren, and all the teachers of the Burke Elementary School, Fairview District, Boulder, Colorado, without whose assistance this experiment could not have been performed. The writers wish to express appreciation to Miss Colette C. Sanville, graduate student, for her assistance in collecting the data of this study.

Child Develpm., 1961, 32, 789-793.

association between stimulus and response components in the learning situation. Hulicka (5) employed a similar definition of the associative variable. Also, in an experiment which requires *Ss* to depress a response button, it is believed that the *speed* with which they *push* the button can be conceived of as an index of *Ss*' incentive to perform the task. Employing these two latencies, the present study investigated the relative effects of immediate and delayed reinforcement on associative and incentive factors in two-stimulus size and form discrimination learning involving children.

METHOD

Subjects

The *Ss* were 20 children: 10 kindergartners, four girls and six boys, and 10 first graders, five girls and five boys. *Ss* were chosen at random from among 46 first graders and 26 kindergartners. The school from which *Ss* were drawn consists of children predominantly from middle class homes.

Materials

The basic materials used in this experiment have been described in detail elsewhere (11). The stimuli were two pairs of three-dimensional geometric objects: for task 1, the size problem, two cube-boxes 2 and 4 sq. in., respectively; for task 2, a form problem, a sphere and a pyramid. The apparatus consisted of a background board with two push-button mounts 6 mm. in height and a panel board locked onto the rear of and at right angles to the background board. On each trial the stimulus set was placed onto the background board immediately in back of the push-button mounts. A reinforcement light was mounted on the panel board. The light went on immediately following *Ss*' responses for the immediate condition and after a seven-second delay for the delayed reinforcement treatment. In order that measurements of association and incentive could be obtained, Standard electric and Hunter electronic timers were added to the basic apparatus. Mounted under the front edge of the background board was a wooden bar containing a slot which housed a microswitch and a 13.5 by 13 in. plywood shield. The Standard and Hunter timers measured reaction time to 1/100 and 1/1000 of a second, respectively. The circuits were so arranged that removal of the shield activated the microswitch which in turn activated the Standard timer. A 1 mm. depression of either of the push buttons broke the circuit to the Standard timer and simultaneously activated the Hunter timer. The Hunter timer was deactivated when the stem of the push button was depressed its full distance by *S*. Measurements of associative and incentive variables were obtained on the Standard and Hunter timers, respectively.

Design and Procedure

Ss were randomly assigned to two groups of 10 children, each containing five kindergartners and five first graders. Group I *Ss* were given the size task under the delayed reinforcement condition and the form task under the immediate reward condition. For group II *S*s, the task-reinforcement combinations were the reverse of those for group I. *S*s, tested individually, received the following instructions: "This is a game where we play with these things (*E* points to the four stimulus objects). You are supposed to choose one of the two things up here (*E* points to the background board) and push the button in front of the one you choose. If you choose the correct one, the light will go on. If you choose the wrong one, the light will not go on." The experimenter informed *S* that the shield would be inserted in place before each trial. This made it impossible for *S* to observe *E* exchange stimulus sets between trials. Also, removal of the shield activated the Standard timer. The delay involved in replacing the shield in the slot after each trial was consistent with the reward condition of the just completed trial. Training was continued until *S*s reached criterion (six consecutive correct responses) on each of the two tasks. For task 1 the correct response was a button-pushing response to the larger of two cube boxes, and for task 2, a response to a sphere instead of a pyramid.

RESULTS

As was predicted, *S*s learned significantly more quickly under the immediate than under the delayed reinforcement condition ($p = .005$). The Wilcoxon matched-pairs signed-ranks test was used to analyze both latency measures (*see* Table 1). The effects of the reinforcement conditions on both measures were determined at three points in the learning process: (a) the first six trials, (b) at the point where *S*s reached criterion on the first of the two tasks (the six adjusted trials),[2] and (c) the six criterion trials. For both of the latency measures a raw score for each *S* for each of the two reinforcement categories was determined by summing the latencies of six trials at the three points in the learning process. The associative latencies, with the exception of the first six trials analysis, were significantly shorter under the delayed than under the immediate condition ($p = .44$, .003, and .02 for the first six trials, adjusted trials, and criterion trials, respectively). The reinforcement conditions did not differentially affect the incentive latencies ($p = .50$, .35, and .62 for the first six trials, adjusted trials, and criterion trials, respectively). It is believed that the incentive measure was not sensitive enough to pick up possible differences in the effect of the treatment conditions.

---

[2] Adjusted scores were determined by the number of trials to the first criterion and the reinforcement condition under which *S*s first reached criterion. For example, if *S* required 10 trials (exclusive of the six criterial trials) under the immediate and 18 trials (exclusive of the six criterial trials) under the delayed reinforcement condition to learn the correct response, the sum of the six criterial trials (trials 11 to 16, inclusive) under the immediate condition was compared with the sum of trials 11 to 16, inclusive, under the delayed reinforcement condition. If *S* reached criterion first under the delayed reinforcement condition, measurements under the immediate condition were adjusted. This adjustment was made for each *S*.

TABLE I

WILCOXON MATCHED-PAIRS SIGNED-RANKS ANALYSIS OF ASSOCIATIVE AND INCENTIVE MEASURES

(For each comparison combination, $N = 20$)

Measure	T	p*
Associative		
First Six Trials	84.5	.44
Adjusted Trials	24.0	.003
Criterion Trials	43.0	.02
Incentive		
First Six Trials	105.0	.50
Adjusted Trials	80.0	.35
Criterion Trials	118.0	.62

* Probabilities are for two-tailed tests of significance.

DISCUSSION

Contrary to impressions suggested by casual observations in a previous experiment (11), *S*s in the present experiment responded more quickly to the task with which delay of reinforcement was associated (even though they learned this less quickly) than to the task with which immediate reinforcement was associated.

The interference of reinforcement delay with the establishment of correct associations may be interpreted in at least two ways: Perhaps delay of reward thwarts problem-solving behavior. The thwarting of goal-directed behavior is frustrating. Numerous studies both at the human and the infrahuman level have shown that frustration (a) impairs performance on relatively complex tasks and (b) is associated with increased muscular tension which produces faster reactions (1, 2, 3, 4, 6). These studies indicate that frustration and its accompanying tension are negatively related to performance. It seems reasonable to suggest that *S*s, in the present experiment, frustrated because of the delay of an indication of the correctness or incorrectness of their response, (a) reacted in an impulsive, nonreflective way, (b) gave little if any attention to the significant elements of the task at hand, and (c) made use of previous experience with the task in an inefficient way. On the other hand, when reinforcement followed correct responses immediately, *S*s were apparently more attentive to the meaningful components of the task situation, a fact which led to a quicker strengthening of the correct habits.

Another possible explanation of these results consists of the likelihood that *S*s became uninterested and consequently bored with a task which involved delay of information concerning the correctness or incorrectness of their responses. Within the framework of this hypothesis, *Ss*' faster reaction time when learning under the delayed reward condition may imply task avoidance. This attitude would also be expected to result in slower learning under the delay condition. The fact that reinforcement conditions affected associative reaction time significantly in the later but not the initial stages of learning lends partial support to this hypothesis.

SUMMARY

This paper reports the results of an investigation of the relative effects of immediate and delayed reinforcements on associative and motivational factors in size and form discrimination learning with children. Twenty *Ss* were used: 10 kindergartners and 10 first graders. Association only was differentially affected by the time interval separating reaction from reinforcement. Frustration-tension and boredom hypotheses were offered in an effort to account for the detrimental effect of delayed reinforcement on association.

REFERENCES

1. BARKER, R. G., DEMBO, T., & LEWIN, K. Frustration and regression: an experiment with young children. *Univer. Iowa Stud. Child Welf.*, 1941, 18, No. 1.
2. DANIEL, R. S. The distribution of muscular action potentials during maze learning. *J. exp. Psychol.*, 1939, 24, 621-629.
3. DAVIS, R. C. Set and muscular tension. *Ind. Univer. Publ. Sci. Ser.*, 1940, No. 10.
4. FREEMAN, G. L. Changes in tonus during completed and interrupted mental work. *J. gen. Psychol.*, 1930, 4, 309-334.
5. HULICKA, I. M. Additive vs. multiplicative combination of drive and incentive. *Psychol. Rep.*, 1960, 6, 403-409.
6. JOST, H. Some physiological changes during frustration. *Child Develpm.*, 1941, 12, 9-15.
7. LIPSITT, L., & CASTANEDA, A. Effects of delayed reward on choice behavior and response speeds in children. *J. comp. physiol. Psychol.*, 1958, 51, 65-67.
8. LOGAN, F. A. The role of delay of reinforcement in determining reaction potential. *J. exp. Psychol.*, 1952, 43, 393-399.
9. PERIN, C. T. The effects of delayed reinforcement upon the differentiation of bar responses in white rats. *J. exp. Psychol.*, 1943, 32, 95-109.
10. SPENCE, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, 54, 1-8.
11. TERRELL, G., & WARE, R. Role of delay of reward in speed of size and form discrimination learning in childhood. *Child Develpm.*, 1961, 32, 409-415.

EAU OF EDNL.

COML. & PSYL.
LIBRARY
21 JAN 1966